

us-09-854-133-586.rag

Wed Jul 9 15:44:27 2003

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: July 9, 2003, 15:05:28 ; Search time 54.0796 Seconds  
(without alignments)  
239.005 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues 908470  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	18.4	16	AAE13851	Human T cell epitope
2	82	15.4	121	AAO13020	Human polypeptide
3	71.5	13.4	46	AAO04640	Human polypeptide
4	71.5	13.4	141	AAO09875	Human polypeptide
5	71.5	13.4	163	AAO09875	Human colon cancer
6	69.5	13.1	66	ABO3142	Human musculoskele
7	69	13.0	63	ABO17957	Human nervous syst
8	68	12.8	423	AAM51507	Mouse 5HT receptor
9	68	12.8	423	AAE03740	Mouse G-protein co
10	68	12.8	423	AAE20598	Homo sapiens gluco

11	68	12.8	458	20	AAV28851	Active form of hum
12	68	12.8	458	21	AAV90676	Human mutant G pro
13	68	12.8	458	22	AAB45803	Human serotonin re
14	67.5	12.7	105	22	AAB93643	Human protein sequ
15	67	12.6	81	23	ABP00516	Human OREF protein
16	67	12.6	190	22	AAB64601	Human secreted pro
17	67	12.6	458	17	AAR94577	Human serotonin va
18	67	12.6	458	19	AAW77108	Human 5-HT2C serot
19	67	12.6	458	21	AAV90641	Human G protein-co
20	66.5	12.5	64	22	AAV95632	Human reproductive
21	66.5	12.5	116	22	AAO12091	Human polypeptide
22	66.5	12.5	138	22	AAO00532	Human polypeptide
23	66.5	12.5	2197	21	AAO00532	Plasmodium falcipa
24	66	12.4	83	22	ABG26719	Novel human diagno
25	66	12.4	100	22	ABG11242	Novel human diagno
26	66	12.4	369	23	AAV51135	Common bean COK-4
27	66	12.4	478	20	AAV28853	Active form of hum
28	66	12.4	478	22	AAV45805	Human serotonin re
29	66	12.4	637	20	AAV13445	Invertebrate octop
30	66	12.4	637	22	ABG60216	Drosophila melanog
31	66	12.4	642	20	AAV33680	D. melanogaster oc
32	66	12.4	642	20	AAV25464	Drosophila melanog
33	66	12.4	738	22	ABB65092	Drosophila melanog
34	65.5	12.3	381	22	AAU45320	Propionibacterium
35	65	12.2	232	21	AAU07926	Arabidopsis thalia
36	65	12.2	418	10	AAV90549	Serotonin 5HT1c re
37	65	12.2	514	20	AAV05896	Vicia sativa fatty
38	65	12.2	1030	22	ABB60070	Drosophila melanog
39	64.5	12.1	59	22	AAV86657	Human immune/haema
40	64.5	12.1	173	23	ABB72349	Murine protein iso
41	64.5	12.1	177	21	AAV47017	Arabidopsis thalia
42	64.5	12.1	184	21	AAV47016	Arabidopsis thalia
43	64.5	12.1	189	22	AAU27511	Human G-protein co
44	64.5	12.1	213	21	AAV14923	Arabidopsis thalia
45	64.5	12.1	213	21	AAV47015	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAE13851  
ID AAE13851 standard; peptide; 16 AA.  
XX AAE13851;  
AC AAE13851;  
XX 26-FEB-2002 (first entry)  
DT Human T cell epitope related to lung tumour-specific protein.  
XX Human; lung tumour protein; immunostimulant; gene therapy;  
DE antisenese-therapy; vaccine; immune response; lung cancer;  
KW T cell epitope.  
KW Homo sapiens.  
OS WO200172295-A2.  
XX 04-OCT-2001.  
XX 28-MAR-2001; 2001WO-US09991.  
XX 29-MAR-2000; 2000US-0538037.  
PR 05-JUN-2000; 2000US-0588937.  
PR 18-AUG-2000; 2000US-0640878.  
PR 22-SEP-2000; 2000US-234517P.  
PR 01-NOV-2000; 2000US-0704512.  
PR 14-DEC-2000; 2000US-0738973.  
XX (CORI-) CORIXA CORP.  
XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
PI

XX DR WPI; 2001-639201/73.  
XX PT New human lung-specific polynucleotides and polypeptides for the  
XX PT diagnosis and treatment of disease e.g. lung cancer -  
XX PS Claim 2; Page 378; 378pp; English.  
XX CC The invention relates to isolated lung tumour-specific proteins and  
CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
CC their antigen-presenting cells are useful for stimulating and/or  
CC expanding T cells specific for a tumour protein, and for inhibiting  
CC the development of cancer. The invention also relates to a composition  
CC useful for stimulating an immune response, and for treating cancer. The  
CC lung tumour specific oligonucleotide is useful in gene therapy and for  
CC diagnosis, detection and treatment of lung cancer. The present sequence  
XX is human T cell epitope related to lung tumour-specific protein.  
SQ Sequence 16 AA;

Query Match 18.4%; Score 98; DB 22; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.0002;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 FQANGCIDFIIFWIFW 50  
Db 1 FQANGCIDFIIFWIFW 16

RESULT 2  
AAO13020  
ID AAO13020 standard; Protein; 121 AA.  
XX  
AC AAO13020;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 26912.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI; 2001-514838/56.  
DR N-PSDB; AAI92951.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 26912; 1399pp + Sequence Listing; English.  
XX

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 121 AA;

Query Match 15.4%; Score 82; DB 22; Length 121;  
Best Local Similarity 34.1%; Pred. No. 0.13;  
Matches 28; Conservative 4; Mismatches 20; Indels 30; Gaps 5;

QY 10 SLGDSETLSQTEL---RKKERKKRERKFOANGC-IDFIIFWIFWIFSHHWHQESLLC 65  
Db 59 SLGDRARLPQKRERERETERKRRERASCAFCGATFVNMW-----CQKHH----- 105  
QY 66 PPSPKVETCREMLTGGCLPWAT 87  
Db 106 -----TCNH-----LPWLT 114

RESULT 3  
AAO04640  
ID AAO04640 standard; Protein; 46 AA.  
XX  
AC AAO04640;  
XX

DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 18532.  
XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI; 2001-514838/56.  
DR N-PSDB; AAI84571.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 18532; 1399pp + Sequence Listing; English.  
XX

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or

us-09-854-133-586.rag

Wed Jul 9 15:44:27 2003

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 46 AA;  
SQ  
Query Match 13.4%; Score 71.5; DB 22; Length 46;  
Best Local Similarity 81.0%; Pred. No. 0.69;  
Matches 17; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 14 SETLSQTELRKKKKRERK 34  
Db 12 SETLSQKE-RKKRKKKKRERK 31

RESULT 4  
AAO09875  
ID AAO09875 standard; Protein; 141 AA.

XX AAO09875;  
XX  
DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 23767.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.  
OS  
XX WO200164835-A2.

PN  
XX 07-SEP-2001.

PD 26-FEB-2001; 2001WO-US04927.

PF 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

PR (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

PI WPI; 2001-514838/56.

XX N-PSDB; AAI89806.

DR Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

PT Claim 20; SEQ ID NO 23767; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
PS the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 141 AA;  
SQ

Query Match 13.4%; Score 71.5; DB 22; Length 141;  
Best Local Similarity 31.0%; Pred. No. 2.5;  
Matches 26; Conservative 13; Mismatches 36; Indels 9; Gaps 3;

QY 1 EVEVSRD--HAGLGDSETLSQTELRKKKKRERK-----RKQANCGIDFIIFWIFWIL 52  
Db 5 KVAVSVDLTMTSHNSLNSSETLSQKEKTKQTKQKKTYXEGRKHKRKCPTPLIIRIQT 64  
QY 53 LFSHHWQI-ESLLCPSPKKEVTCR 75  
Db 65 TVRYHFTIRIVLLPQKQKNCWCR 88

RESULT 5  
AAG75462  
ID AAG75462 standard; Protein; 163 AA.

XX AAG75462;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6226.

DE Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma.  
XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

PI WPI; 2001-235357/24.

XX N-PSDB; AAH34867.

DR Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -

PT Claim 11; Page 7674-7676; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
PS cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing of P.  
CC inactive proteins or to supplement the patients own cancer-associated Ps,  
CC Additionally, N may be used to produce the colon cancer-associated cell  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAG77789 represent sequences used in the exemplification of the  
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 163 AA;  
SQ

Query Match 13.4%; Score 71.5; DB 22; Length 163;  
Best Local Similarity 38.6%; Pred. No. 2.9;

	Qy	Db	Matches	9;	Indels	21;	Gaps	4;
10	SLGD-SETLSOTELRKKRKKRKFQANGCIDFIWFIFWILLFSHHWQESLLC	65						
11	SLGDKSETLSP-----KKKKKKKN-----WIAW--LYSGHSMQAQFCC	47						

RESULT 6  
ABB03142  
ID ABB03142 standard; Protein; 66 AA.  
XX  
AC ABB03142;  
XX  
DT  
XX  
XX 08-JAN-2002 (first entry)  
DE Human musculoskeletal system related polypeptide SEQ ID NO 1089.  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system.

WO200155367-A1.

02-AUG-2001.

17-JAN-2007:

SECRETOR

31-JAN-2000; 2000US-0179065.

24-FEB-2000; 2000US-0184664.

16-MAR-2000; 2000US-0186350.  
17-MAR-2000; 2000US-0189874.

17 - MAR - 2000; 2000US-0190076.  
18 - APR - 2000; 2000US-0198133

19-MAY-2000; 2000US-0205515.  
07-JUN-2000; 2000US-0200467.

8-JUN-2000; 2000US-0214886.  
10-JUN-2000; 2000US-0214887.

7-JUL-2000; 2000US-0216647.

1-JUL-2000; 2000US-0216880;  
1-JUL-2000; 2000US-0217487;

4 - JUL - 2000; 2000US-0218290.

6-JUL-2000; 2000US-0220963.  
6-JUL-2000; 2000US-0220964

4-AUG-2000; 2000US-0224518.  
4-AUG-2000; 2000US-0334518.

1-AUG-2000; 2000US-0225213.  
1-AUG-2000; 2000US-0225213.

1-AUG-2000; 2000US-0225266.

-AUG-2000; 2000US-0225268.

2000US-0225270.  
2000US-0225447

-AUG-2000; 2000US-0225757.  
-AUG-2000; 2000US-0225758

-AUG-2000; 2000US-0225759.  
-AUG-2000; 2000US-0225759.

- AUG-2000; 2000US-0226681.  
- AUG-2000; 2000US-0226681.

-AUG-2000; 2000US-0226868.  
-AUG-2000; 2000US-0227182.

2000US-0227009;  
-AUG-2000; 2000US-0228924;

SEP-2000; 2000US-0229343

SEP-2000; 2000US-0229344.  
SEP-2000; 2000US-0229345

SEP-2000; 2000US-0229509.

\_\_\_\_\_



us-09-854-133-586.rag

Wed Jul 9 15:44:27 2003

PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 17-NOV-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-451937/48.  
DR N-PSDB; AAL34724.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the musculoskeletal system including  
PT musculoskeletal cancers and also for testing and detection e.g.  
PT diagnosis -  
XX  
XX Claim 11; SEQ ID NO 1089; 781pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (AAL34669-AAL37666) and proteins  
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful  
CC for preventing, treating or ameliorating medical conditions e.g. by  
CC protein or gene therapy. The genes are isolated from a range of human  
CC tissues disclosed in the specification. The nucleic acids, proteins,  
CC antibodies and (ant)agonists are useful in the diagnosis, treatment  
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and  
CC other cancers of the adrenal gland, bone, bone marrow, breast,  
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
CC and (f) infectious diseases such as viral, bacterial, fungal and  
CC parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 66 AA;  
SQ  
Query Match 13.1%; Score 69.5; DB 22; Length 66;  
Best Local Similarity 50.0%; Pred. No. 1.8;  
Matches 19; Conservative 5; Mismatches 7; Indels 7; Gaps 2;  
QY 1 EVEVSRDHASL---GDSETLSQTELKRRKKRKRK 34  
Db 5 EAAVSCDHAGVLQPGQHSLSQ---NKKERKRRK 39  
RESULT 7  
ABB17957  
ID ABB17957 standard; Protein; 63 AA.  
XX  
AC ABB17957;  
XX  
DT 23-JAN-2002 (first entry)  
XX Human nervous system related polypeptide SEQ ID NO 6614.  
DE  
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnary;  
KW antiparkinsonian; antispasmodic; antianemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.  
XX  
XX Homo sapiens.  
OS  
XX WO200159063-A2.  
PN  
XX  
XX 16-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01334.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0226279.  
PR 18-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR

PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0242221.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.

PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

N-PSDB; ABA14283.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases.

Claim 11; SEQ ID NO 6614; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABA11004-ABA21534) and proteins (ABA14678-ABB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 63 AA;

Query Match

Best Local Similarity 13.0%; Score 69; DB 22; Length 63;

Matches 18; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

QY 2 VEVSRRDASL---GDSETLSQTELKRRKKRRE 32

Db 6 VAVSQDSATTLPQGRQSKTSLQKKKKKKRKKRRE 40

RESULT 8

AAM51507

ID AAM51507 standard; Protein; 423 AA.

XX AAM51507;

AC AAM51507;

XX 02-JAN-2002 (first entry)

DT Mouse 5HT receptor.

DE Mouse; 5HT receptor; chimera; cDNA library preparation; serotonin;

KW G protein-coupled receptor.

XX Mus musculus.

OS JP2001178476-A.

PN 03-JUL-2001.

XX

PD



```

PR 12-JUL-2000; 2000US-217537P.
PR 12-JUL-2000; 2000US-218069P.
PR 12-JUL-2000; 2000US-218074P.
PR 12-JUL-2000; 2000US-218358P.
PR 27-JUL-2000; 2000US-221483P.
PR 07-AUG-2000; 2000US-223120P.
PR 07-AUG-2000; 2000US-223122P.
PR 26-OCT-2000; 2000US-243958P.
PR 15-NOV-2000; 2000US-249408P.
PR 20-NOV-2000; 2000US-252299P.
PR 16-JAN-2001; 2001US-262113P.
PR 16-JAN-2001; 2001US-262205P.
XX
PA (DELT-) DELTAGEN INC.
XX
XX Allen KD, Brennan TJ;
XX
XX WPI; 2002-164574/21.
XX N-PSDB; AAD32919.
XX
XX Novel non-human transgenic animal, especially transgenic mice useful
XX for identifying an agent that modulates expression or function of
XX target gene, comprises disruptions in target G protein-coupled receptor
XX gene -
XX
XX Disclosure; Fig 7; 101pp; English.
XX
XX The invention relates to a non-human transgenic animal having targeted
XX G-protein coupled receptor (GPCR) gene disruptions in melanocortin-3 gene
XX 5-HT-2B gene, chemokine receptor 9A gene, glucocorticoid-induced receptor
XX gene, orphan GPR10 (UHR)-1 gene, orphan GPR14 gene, orphan GPR15 gene,
XX beta chemokine receptor (E01) gene, endothelial differentiation GPCR3
XX (EDG3) gene, ATP receptor P2U1 gene or adenosine 3 receptor gene. The
XX transgenic animal is useful for identifying an agent that modulates the
XX expression or function of the target gene, for identifying an agent that
XX ameliorates a phenotype associated with a disruption in the target gene.
XX A transgenic construct is useful for producing a transgenic animal,
XX preferably a transgenic mouse. The transgenic animal is useful for
XX testing the efficacy of proposed genetic and pharmacological therapies
XX for human genetic diseases, such as neurological, neuropsychological or
XX psychotic illnesses. The transgenic animal is also useful as models for
XX diseases, disorders or conditions associated with phenotypes relating to
XX a disruption in a target, and to identify pharmaceuticals, therapies,
XX drugs and interventions which may be effective in treating a disease or
XX other phenotypic characteristics of the animal. An agent which modulates
XX the expression of the target gene is useful as a therapeutic for treating
XX conditions associated with a disruption of the target gene. The present
XX sequence is human glucocorticoid-induced receptor protein.
XX
SQ Sequence 423 AA;
Query Match 12.8%; Score 68; DB 23; Length 423;
Best Local Similarity 22.8%; Pred. No. 22;
Matches 29; Conservative 11; Mismatches 41; Indels 46; Gaps 5;
QY 10 SLGDSSETLSQTELKKRKKRKKRKFQANGIDFIIFIWI---FWILL----- 53
Db :||| | || |||| | :| :| :||
271 TIGDVTTEQYIALR---RKKKTTVKMLVLVVLFALCWFPPLNCYVLLSSKAIHTNNALY 327
QY 54 FSHHWIQESLLC-----PPSPKEVTCREMLTGGCLPWAT 87
Db |:||| | |
328 FAFHWFAMSSSTCYNPFYICWLNENFRVELKALLSMCQRPPKQEDRLPSVPVSRVATWE 387
QY 88 RSHLGRR 94
Db :|| |||
388 KSH-GRR 393
RESULT 11
AAY28851
ID AAY28851 standard; Protein; 458 AA.
XX
XX AC AAY28851;

```



```

Db      329 ITNIL--SVLC-----EKSCNQKL 345

RESULT 13
AAB45803
ID      AAB45803 standard; Protein; 458 AA.
XX
AC      AAB45803;
XX
DT      20-MAR-2001 (first entry)
XX
DE      Human serotonin receptor associated protein SEQ ID NO 23.
XX
KW      Human serotonin receptor; 5-HT2A receptor; 5-HT2C receptor; agonist;
KW      1-Phenyl-3-(3-(4-bromo-3-pyrazolyl)phenyl)-urea derivative; modulator
KW      central nervous system disorder; CNS; gastrointestinal disorder;
KW      cardiovascular disorder; inflammatory disorder.
XX
OS      Homo sapiens.
XX
PN      US6150393-A.
XX
PD      21-NOV-2000.
XX
PF      15-OCT-1999; 99US-0418721.
XX
PR      18-DEC-1998; 98US-0112909.
PR      05-MAR-1999; 99US-0123000.
PR      07-SEP-1999; 99US-0152708.
PR      14-APR-1999; 99US-0292069.
PR      14-APR-1999; 99US-0292071.
PR      14-APR-1999; 99US-0292072.
XX
PA      (AREN-) ARENA PHARM INC.
XX
PI      Foster RJ, Chalmers DT, Behan DP, Liu Q, Menzaghi F, Russo JF;
PI      Smith JR, Thomsen WJ, Glen RC, Lawless MS, Liaw CW, Beeley NRA;
XX      WPI; 2001-079410/09.
DR      N-PSDB; AAC82826.
XX
PT      New 1-phenyl-3-(3-(4-bromo-3-pyrazolyl)phenyl)-urea derivatives, use
PT      as 5-HT(2A) receptor ligands e.g. for treating central nervous system
PT      gastrointestinal, cardiovascular and inflammatory disorders -
XX
PS      Disclosure; Column 55-58; 44pp; English.
XX
CC      This invention describes novel 1-Phenyl-3-(3-(4-bromo-3-pyrazolyl)ph)ph
CC      -urea derivatives (I) which are modulators of non-endogenous,
CC      constitutively activated forms of serotonin 5-HT2A (and 5-HT2C) rece
CC      and inverse agonists to 5-HT2A receptors. (I) modulate serotonin 5-HT
CC      receptor activity and are potentially useful for treating central ne
CC      system (CNS), gastrointestinal, cardiovascular and inflammatory
XX      disorders.
SQ      Sequence 458 AA;

Query Match      12.8%; Score 68; DB 22; Length 458;
Best Local Similarity 27.4%; Pred. No. 24;
Matches 23; Conservative 15; Mismatches 28; Indels 18; Gaps

QY      6 RDHASLGDSETLSQTELKKKKKKR-----ERKFOANGCIDFIIFWIFWILL
      I: | | | | | : | : : : | | : | | : | | : | | |
DB      269 RNTABEENSANPNQDNARRRRKKKRRPRGTMQAINNERKAKKVLGIVFFVFLIMWCPE

QY      55 SHHWIQESLLCPPSPKEVTCREML 78
DB      329 ITNIL--SVLC-----EKSCNQKL 345

RESULT 14
AAB93643
ID      AAB93643 standard; Protein; 105 AA.

```



QY 18 SQTELRRKKRRKKRRKFFQANCIDFIIF 46  
| :|||| :|||| : | ||  
Db 11 SHRQLRKKAKKKRRARINTFCVFIFFIF 39

Search completed: July 9, 2003, 15:07:40  
Job time : 56.0796 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 30.0442 Seconds  
(without alignments)  
94.994 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	12.8	458	3 US-09-292-071-29	Sequence 29, Appl
2	68	12.8	458	4 US-09-292-069A-29	Sequence 29, Appl
3	68	12.8	458	4 US-09-767-013-29	Sequence 29, Appl
4	67	12.6	458	1 US-08-310-271-2	Sequence 2, Appl
5	67	12.6	458	3 US-09-292-071-27	Sequence 27, Appl
6	67	12.6	458	4 US-09-292-069A-27	Sequence 27, Appl
7	67	12.6	458	4 US-09-032-742-10	Sequence 10, Appl
8	67	12.6	458	4 US-09-767-013-27	Sequence 27, Appl
9	66	12.4	478	3 US-09-292-071-33	Sequence 33, Appl
10	66	12.4	478	4 US-09-292-069A-33	Sequence 33, Appl
11	66	12.4	478	4 US-09-767-013-33	Sequence 33, Appl
12	64	12.0	380	4 US-09-134-001C-3830	Sequence 3830, Ap
13	64	12.0	488	4 US-09-215-694-15	Sequence 15, Appl
14	62	11.7	478	3 US-09-292-071-31	Sequence 31, Appl
15	62	11.7	478	4 US-09-292-069A-31	Sequence 31, Appl
16	62	11.7	478	4 US-09-767-013-31	Sequence 31, Appl
17	61.5	11.6	111	1 US-08-334-773A-6	Sequence 6, Appl
18	61.5	11.6	365	2 US-08-428-243-9	Sequence 9, Appl
19	61.5	11.6	365	5 PCT-US93-10301-9	Sequence 9, Appl
20	61.5	11.6	422	1 US-07-996-772A-12	Sequence 12, Appl
21	61.5	11.6	445	2 US-08-157-185-2	Sequence 2, Appl
22	61.5	11.6	445	3 US-08-281-526B-2	Sequence 2, Appl
23	61.5	11.6	445	4 US-09-450-797-2	Sequence 2, Appl
24	61.5	11.6	445	4 US-09-328-314-16	Sequence 16, Appl
25	61.5	11.6	445	4 US-09-450-790A-2	Sequence 2, Appl
26	61.5	11.6	445	4 US-09-332-837-2	Sequence 2, Appl
27	61.5	11.6	445	5 PCT-US93-10553-2	Sequence 2, Appl

28	61	11.5	62	4 US-08-924-629C-71	Sequence 71, Appl
29	60.5	11.4	468	4 US-09-092-636-9	Sequence 9, Appl
30	60	11.3	460	4 US-09-032-742-20	Sequence 20, Appl
31	59.5	11.2	404	2 US-08-428-243-7	Sequence 7, Appl
32	59.5	11.2	404	5 PCT-US93-10301-7	Sequence 7, Appl
33	59.5	11.2	435	2 US-08-031-538-11	Sequence 11, Appl
34	59.5	11.2	2441	1 US-08-194-468-2	Sequence 2, Appl
35	59.5	11.2	2441	3 US-08-961-739-2	Sequence 2, Appl
36	59.5	11.2	2441	4 US-09-514-247A-8	Sequence 8, Appl
37	59.5	11.2	2442	4 US-09-514-247A-10	Sequence 10, Appl
38	59	11.1	160	2 US-08-726-306A-64	Sequence 64, Appl
39	59	11.1	349	1 US-08-118-270-71	Sequence 71, Appl
40	59	11.1	349	5 PCT-US93-08528-71	Sequence 71, Appl
41	59	11.1	385	1 US-08-118-270-31	Sequence 31, Appl
42	59	11.1	385	5 PCT-US93-08528-31	Sequence 31, Appl
43	59	11.1	453	1 US-08-313-553-11	Sequence 11, Appl
44	59	11.1	453	3 US-08-767-993-11	Sequence 11, Appl
45	59	11.1	460	1 US-07-817-920-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-09-292-071-29  
; Sequence 29, Application US/09292071  
; Patent No. 6107324  
; GENERAL INFORMATION:  
; APPLICANT: Behan, Dominic  
; APPLICANT: Chalmers, Derrick  
; TITLE OF INVENTION: No. 6107324-Endogenous, Constitutively Activated  
; TITLE OF INVENTION: Human Serotonin Receptors and Small Molecule Modulators  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arena Pharmaceuticals, Inc.  
; STREET: 6166 Nancy Ridge Drive  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/292,071  
; FILING DATE: April 14, 1999  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Rosen  
; REGISTRATION NUMBER: 39,822  
; TELEPHONE: (215) 564-6525  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 29:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 458 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
US-09-292-071-29

Query Match 12.8%; Score 68; DB 3; Length 458;  
Best Local Similarity 27.4%; Pred. No. 5.5;  
Matches 23; Conservative 15; Mismatches 28; Indels 18; Gaps 3;

QY	6	RDHASLGDSSETLSQTELRKKRKKR-----ERKFAQNCGIDFIWFIFWILLF	54
Db	269	RNTAEENSANPNQDNARRRRKKRRPRGTMQAINNERKAKKVLGIVFFVFLIMWCPEFF	328
QY	55	SHHWIQESLLCPPSPKEVTCREML	78

4

1

---

RESULT 5  
US-09-292-071-27  
; Sequence 27, Application US/09292071  
; Patent No. 6107324  
; GENERAL INFORMATION:  
; APPLICANT: Behan, Dominic  
; APPLICANT: Chalmers, Derick  
; TITLE OF INVENTION: No. 6107324-Endogenous, Constitutively Activated  
; TITLE OF INVENTION: Human Serotonin Receptors and Small Molecule Modulators Thereof  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arena Pharmaceuticals, Inc.  
; STREET: 6166 Nancy Ridge Drive  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/292,071  
; FILING DATE: April 14, 1999  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Rosen  
; REGISTRATION NUMBER: 39,822  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 564-6525  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 458 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
US-09-292-071-27

Query Match 12.6%; Score 67; DB 3; Length 458;  
Best Local Similarity 27.4%; Pred. No. 7.2;  
Matches 23; Conservative 14; Mismatches 29; Indels 18; Gaps 3;

QY 6 RDHASLGDSFTLSQTSLRKKKR-----ERKFAQNCGIDFIIFWIFWILFF 54  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 269 RNTAEENSANPNQDNARRRRKKRRPRGTMQAINNERKASKVLGIIVFFVFLIMWCFFF 328  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 55 SHHWIQESLLCPPSPKEVTCREML 78  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 329 ITNIL--SVLC-----EKSCNQKL 345

RESULT 6  
US-09-292-069A-27  
; Sequence 27, Application US/09292069A  
; Patent No. 6140509  
; GENERAL INFORMATION:  
; APPLICANT: Behan, Dominic P  
; APPLICANT: Chalmers, Derek T  
; APPLICANT: Foster, Richard J

RESULT 7  
US-09-032-742-10  
; Sequence 10, Application US/09032742  
; Patent No. 6255089  
; GENERAL INFORMATION:  
; APPLICANT: Teitler, Milt  
; APPLICANT: Herrick-Davis, Katharine  
; APPLICANT: Egan, Christina C.  
; TITLE OF INVENTION: Constitutively Activated Serotonin  
; TITLE OF INVENTION: Receptors  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Laurence Weinberger  
; STREET: 882 S. Matlack Street, Suite 103  
; STREET: P.O. Box 1663  
; CITY: West Chester  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19380-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/032,742  
; FILING DATE: 27-FEB-1998  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weinberger, Laurence

\_\_\_\_\_





COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/292,071  
FILING DATE: April 14, 1999  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Rosen  
REGISTRATION NUMBER: 39,822  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 564-6525  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 478 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-09-292-071-31

Query Match 11.7%; Score 62; DB 3; Length 478;  
Best Local Similarity 24.7%; Pred. No. 29;  
Matches 22; Conservative 16; Mismatches 33; Indels 18; Gaps 3;  
QY 6 RDHASLGDSETLSQTELRRKKKKR-----ERKFAQNCGIDFIIFWIFWILLF 54  
Db 289 RNTAEENSANPNQDNARRRRKKRRPRGTMQAINNERKASKVLGIVFFLVVMWCPFF 348  
QY 55 SHHWIQESLLCPPSPKVTCREMLTGGCL 83  
Db 349 ITNIM--AVICKES-----CNEDVIGALL 370

RESULT 15  
US-09-292-069A-31  
Sequence 31, Application US/09292069A  
Patent No. 6140509  
GENERAL INFORMATION:  
APPLICANT: Behan, Dominic P  
APPLICANT: Chalmers, Derek T  
APPLICANT: Foster, Richard J  
APPLICANT: Glen, Robert C  
APPLICANT: Lawless, Michael S  
APPLICANT: Liaw, Chen W  
APPLICANT: Liu, Qian  
APPLICANT: Russo, Joseph F  
APPLICANT: Smith, Julian R  
APPLICANT: Thomsen, William J  
TITLE OF INVENTION: No. 6140509-Endogenous, Constitutively Activated Human  
TITLE OF INVENTION: Serotonin Receptors And Small Molecule Modulators  
FILE REFERENCE: AREN0033  
CURRENT APPLICATION NUMBER: US/09/292,069A  
CURRENT FILING DATE: 1999-04-14  
PRIOR APPLICATION NUMBER: 60/090,783  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/112,909  
PRIOR FILING DATE: 1998-12-18  
PRIOR APPLICATION NUMBER: 60/123,000  
PRIOR FILING DATE: 1999-03-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 31  
LENGTH: 478  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: No. 6140509el  
OTHER INFORMATION: Sequence  
US-09-292-069A-31

Query Match 11.7%; Score 62; DB 4; Length 478;  
Best Local Similarity 24.7%; Pred. No. 29;  
Matches 22; Conservative 16; Mismatches 33; Indels 18; Gaps 3;  
QY 6 RDHASLGDSETLSQTELRRKKKKR-----ERKFAQNCGIDFIIFWIFWILLF 54  
Db 289 RNTAEENSANPNQDNARRRRKKRRPRGTMQAINNERKASKVLGIVFFLVVMWCPFF 348  
QY 55 SHHWIQESLLCPPSPKVTCREMLTGGCL 83  
Db 349 ITNIM--AVICKES-----CNEDVIGALL 370

Search completed: July 9, 2003, 15:08:26  
Job time : 35.0443 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:06:38 ; Search time 32.6195 Seconds  
(without alignments)  
346.196 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues  
Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US60\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	532	100.0	97	9 US-09-854-133-586	Sequence 586, App
2	532	100.0	97	9 US-10-144-649A-586	Sequence 586, App
3	532	100.0	97	10 US-09-738-973-586	Sequence 586, App
4	532	100.0	114	9 US-10-144-649A-742	Sequence 742, App
5	98	18.4	16	9 US-09-854-133-587	Sequence 587, App
6	98	18.4	16	9 US-10-144-649A-587	Sequence 587, App
7	98	18.4	16	10 US-09-738-973-587	Sequence 587, App
8	80.5	15.1	119	9 US-09-892-877-400	Sequence 400, App
9	80.5	15.1	119	9 US-09-948-783-400	Sequence 400, App
10	71.5	13.4	163	9 US-10-106-698-6236	Sequence 6236, Ap
11	69.5	13.1	66	10 US-09-764-877-1089	Sequence 1089, Ap
12	68	12.8	423	9 US-09-903-396A-2	Sequence 2, Appli
13	68	12.8	458	9 US-10-251-385-230	Sequence 230, App
14	67	12.6	458	9 US-10-251-385-126	Sequence 126, App
15	67	12.6	458	9 US-10-225-567A-16	Sequence 16, Appl
16	67	12.6	1601	10 US-09-862-027-40	Sequence 40, Appl
17	66.5	12.5	64	9 US-09-764-891-4290	Sequence 4290, Ap
18	66	12.4	369	10 US-09-838-955-3	Sequence 3, Appli
19	66	12.4	673	9 US-10-157-031-291	Sequence 291, App

20	64.5	12.1	173	9 US-09-866-050A-673	Sequence 673, App
21	64	12.0	487	10 US-09-801-368-24	Sequence 24, Appl
22	63.5	11.9	53	9 US-10-001-857-155	Sequence 155, App
23	62.5	11.7	150	10 US-09-864-761-38925	Sequence 38925, A
24	61.5	11.6	67	10 US-09-864-761-37546	Sequence 37546, A
25	61.5	11.6	432	9 US-09-877-843-41	Sequence 41, Appl
26	61.5	11.6	445	9 US-09-877-843-39	Sequence 39, Appl
27	61.5	11.6	445	9 US-10-225-567A-22	Sequence 22, Appl
28	61.5	11.6	445	10 US-09-989-861-16	Sequence 16, Appl
29	61.5	11.6	479	9 US-09-877-843-40	Sequence 40, Appl
30	61	11.5	62	9 US-09-883-343A-71	Sequence 71, Appl
31	61	11.5	641	9 US-10-242-943-4	Sequence 4, Appli
32	60.5	11.4	309	9 US-10-060-795B-3	Sequence 3, Appli
33	60.5	11.4	448	9 US-09-877-843-43	Sequence 43, Appl
34	60	11.3	451	10 US-09-764-864-1124	Sequence 1124, Ap
35	60	11.3	459	9 US-10-280-858-1	Sequence 1, Appli
36	60	11.3	459	9 US-10-280-858-3	Sequence 3, Appli
37	60	11.3	459	9 US-10-280-858-5	Sequence 5, Appli
38	60	11.3	459	9 US-10-280-858-7	Sequence 7, Appli
39	60	11.3	459	9 US-10-280-858-9	Sequence 9, Appli
40	60	11.3	459	9 US-10-280-858-11	Sequence 11, Appl
41	60	11.3	459	9 US-10-280-858-13	Sequence 13, Appl
42	60	11.3	459	9 US-10-280-858-17	Sequence 17, Appl
43	60	11.3	459	9 US-10-280-858-19	Sequence 19, Appl
44	60	11.3	459	9 US-10-280-858-21	Sequence 21, Appl
45	60	11.3	459	9 US-10-280-858-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1  
US-09-854-133-586  
; Sequence 586, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 586  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-586

Query Match	100.0%	Score 532;	DB 9;	Length 97;
Best Local Similarity	100.0%	Pred. No. 1e-48;		
Matches	97;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
Qy	1	EVEVSRDHASLGDSSETLSQTELKRRKKRERKFKQANC	GIDFIIFWIFWILLFSHHW	IQ 60
Db	1	EVEVSRDHASLGDSSETLSQTELKRRKKRERKFKQANC	GIDFIIFWIFWILLFSHHW	IQ 60
Qy	61	ESLLCPPSPKEVTCREMLTGGCLPWATRSHLGRKCS	97	
Db	61	ESLLCPPSPKEVTCREMLTGGCLPWATRSHLGRKCS	97	

RESULT 2  
US-10-144-649A-586  
; Sequence 586, Application US/10144649A  
; Publication No. US20030118599A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.

```
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-586
```

```
Query Match      100.0%; Score 532; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 1e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVEVSRDHASLGDSSETLSQTELKRRKKRKFQANCIDFIIFWIFWILLFHHWQ 60
Db 1 EVEVSRDHASLGDSSETLSQTELKRRKKRKFQANCIDFIIFWIFWILLFHHWQ 60

QY 61 ELLCPPSPKPVTCREMLTGGCLPWATRSHLGRKCS 97
Db 61 ELLCPPSPKPVTCREMLTGGCLPWATRSHLGRKCS 97
```

```
RESULT 3
US-09-738-973-586
; Sequence 586, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 586
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-586
```

```
Query Match      100.0%; Score 532; DB 10; Length 97;
Best Local Similarity 100.0%; Pred. No. 1e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVEVSRDHASLGDSSETLSQTELKRRKKRKFQANCIDFIIFWIFWILLFHHWQ 60
Db 1 EVEVSRDHASLGDSSETLSQTELKRRKKRKFQANCIDFIIFWIFWILLFHHWQ 60

QY 61 ELLCPPSPKPVTCREMLTGGCLPWATRSHLGRKCS 97
Db 61 ELLCPPSPKPVTCREMLTGGCLPWATRSHLGRKCS 97
```

```
RESULT 4
US-10-144-649A-742
; Sequence 742, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144,649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 742
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-742
```

```
Query Match      100.0%; Score 532; DB 9; Length 114;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVEVSRDHASLGDSSETLSQTELKRRKKRKFQANCIDFIIFWIFWILLFHHWQ 60
Db 18 EVEVSRDHASLGDSSETLSQTELKRRKKRKFQANCIDFIIFWIFWILLFHHWQ 77

QY 61 ELLCPPSPKPVTCREMLTGGCLPWATRSHLGRKCS 97
Db 78 ELLCPPSPKPVTCREMLTGGCLPWATRSHLGRKCS 114
```

```
RESULT 5
US-09-854-133-587
; Sequence 587, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 587
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-587
```

```
Query Match      18.4%; Score 98; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00064;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 FOANCIDFIIFWIFW 50
Db 1 FOANCIDFIIFWIFW 16
```

```
RESULT 6
US-10-144-649A-587
; Sequence 587, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
```





\_\_\_\_\_

10

100













R;Claverie, J.M.

personal communication, 1992

A;Reference number: A40201

A;Accession: F40201

A;Molecule type: DNA

A;Residues: 1-673 <CLA>

R;Claverie, J.M.

Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search: Alu-derived and other potential

A;Reference number: A40200; MUID:92241891; PMID:1572661

A;Contents: annotation

C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'x'.

C;Comment: Any significant similarity of a predicted protein sequence to a portion of th

#### Query Match

Best Local Similarity 12.4%; Score 66; DB 4; Length 673;

Matches 19; Conservative 17; Mismatches 31; Indels 22; Gaps 3;

QY

9 ASLGDSETLSQTELRKKRKKRKKRKFQANCIDFIIFWIFWILLFSSHWHWIOE 61

Db

310 AGVSEQGPISKTTKTKKQKKKKKPQKXXFC-----FWGFFFFFCFVFVEIGPCSDT 363

QY

69 PKEVTCREMLTGGCLPWATRSHLGRKCS 97

Db

364 PAGV-----QWQVLAH-----CS 376

#### RESULT 12

A64915

ycdy protein homolog bl591 - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Mar-2002

C;Accession: A64915

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: A64915

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-207 <BLAT>

A;Cross-references: GB:AE000254; GB:U00096; NID:g1787862; PIDN:AAC74663.1; PID:g1787874;

A;Experimental source: strain K-12, substrain MGL655

C;Superfamily: Escherichia coli ycdy protein

#### Query Match

Best Local Similarity 12.3%; Score 65.5; DB 1; Length 207;

Matches 25; Conservative 14; Mismatches 33; Indels 15; Gaps 4;

QY

2 VEVSRDHASLGDSETLSQTELRKKRKKRKKRKFQANCIDFIIFWIFWILLFSSHWHWIOE 61

Db

93 VWLDRESVLFGDS-TLA---LRQWMREKGIQFEMKQNEPEDH-----FGSLLMLAAWLAE 143

QY

62 SLLCPPSPKVEVTCREMLTGGCLPWATR 88

Db

144 -----NGRQTECEELLAHLFPWSTR 164

#### RESULT 13

A90916

probable oxidoreductase component ECs2297 [imported] - Escherichia coli (strain O157:H7,

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C;Accession: A90916

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: A90916

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-207 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA035720.1; PID:g13361763; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs2297

C;Superfamily: Escherichia coli ycdy protein

#### Query Match

Best Local Similarity 12.3%; Score 65.5; DB 2; Length 207;

Matches 25; Conservative 14; Mismatches 33; Indels 15; Gaps 4;

QY

2 VEVSRDHASLGDSETLSQTELRKKRKKRKKRKFQANCIDFIIFWIFWILLFSSHWHWIOE 61

Db

93 VWLDRESVLFGDS-TLA---LRQWMREKGIQFEMKQNEPEDH-----FGSLLMLAAWLAE 143

QY

62 SLLCPPSPKVEVTCREMLTGGCLPWATR 88

Db

144 -----NGRQTECEELLAHLFPWSTR 164

#### RESULT 14

F85764

probable oxidoreductase component Z2581 [imported] - Escherichia coli (strain O157:

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C;Accession: F85764

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; M

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apo

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: F85764

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-207 <STO>

A;Cross-references: GB:AE005174; NID:g12515566; PIDN:AAG56578.1; GSPDB:GN00145; UWG

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: Z2581

C;Superfamily: Escherichia coli ycdy protein

#### Query Match

Best Local Similarity 12.3%; Score 65.5; DB 2; Length 207;

Matches 25; Conservative 14; Mismatches 33; Indels 15; Gaps 4;

QY

2 VEVSRDHASLGDSETLSQTELRKKRKKRKKRKFQANCIDFIIFWIFWILLFSSHWHWIOE 61

Db

93 VWLDRESVLFGDS-TLA---LRQWMREKGIQFEMKQNEPEDH-----FGSLLMLAAWLAE 143

QY

62 SLLCPPSPKVEVTCREMLTGGCLPWATR 88

Db

144 -----NGRQTECEELLAHLFPWSTR 164

#### RESULT 15

T08014

cytochrome P450 CYP94A1 - spring vetch

C;Species: Vicia sativa (spring vetch, tare)

C;Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 16-Feb-2001

C;Accession: T08014

R;Tijet, N.; Helvig, C.; Pinot, F.; le Bouquin, R.; Lesot, A.; Durst, F.; Salaun, J.

Biochem. J. 332, 583-589, 1998

A;Title: Functional expression in yeast and characterization of a clofibrate-inducib

A;Reference number: Z16287; MUID:98364856; PMID:9601090

A;Accession: T08014

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-514 <TIJ>

A;Cross-references: EMBL:AF030260; NID:g4204094; PIDN:AAD10204.1; PID:g4204095

C;Genetics:

A;Gene: vagh111



Search completed: July 9, 2003, 15:09:44  
Job time : 46.7788 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 12.0177 Seconds  
(without alignments)  
334.773 Million cell updates/sec

Title: US-09-854-133-586  
Perfect score: 532  
Sequence: 1 EVEVSRDHASLGDSSETLSQT.....LTGGCLPWATRSHLGRKCS 97

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	68	12.8	423	1 GP72_MOUSE	P30731 mus musculu
2	67	12.6	458	1 5H2C_HUMAN	P28335 homo sapien
3	65.5	12.3	207	1 YNFI_ECOLI	P76174 escherichia
4	65	12.2	514	1 C941_VICSA	O81117 vicia sativ
5	63	11.8	353	1 APJ_XENLA	P79960 xenopus lae
6	62.5	11.7	177	1 ATPD_VIBAL	P12987 vibrio algi
7	62.5	11.7	728	1 BPHY_PSEAE	Q9hwr3 pseudomonas
8	62	11.7	1549	1 TRHY_SHEEP	P22793 ovis aries
9	61.5	11.6	218	1 ACDL_CLOPE	Q8xmy2 clostridium
10	61.5	11.6	479	1 5H7_HUMAN	P34969 homo sapien
11	61.5	11.6	568	1 RSL_RHIME	P14129 rhizobium m
12	61.5	11.6	2164	1 POLG_HRV89	P07210 human rhino
13	61	11.5	62	1 PPAL_PEDAC	P29430 pediococcus
14	60.5	11.4	468	1 ETS2_MOUSE	P15037 mus musculu
15	60	11.3	108	1 VE4_HPV11	P04016 human papil
16	60	11.3	108	1 VE4_HPV6C	P20969 human papil
17	60	11.3	459	1 5H2C_MOUSE	P34968 mus musculu
18	60	11.3	483	1 CBF5_YEAST	P33322 saccharomyc
19	59.5	11.2	227	1 MYOG_CHICK	P17920 gallus gall
20	59.5	11.2	333	1 ANR2_HUMAN	Q9qzv1 homo sapien
21	59.5	11.2	446	1 5H7_CAVPO	P50407 cavia porce
22	59.5	11.2	448	1 5H7_MOUSE	P32304 mus musculu
23	59.5	11.2	448	1 5H7_RAT	P32305 rattus norv
24	59.5	11.2	487	1 HH1R_HUMAN	P35367 homo sapien
25	59.5	11.2	639	1 ACM3_CHICK	P49578 gallus gall
26	59.5	11.2	2441	1 CBP_MOUSE	P45481 mus musculu
27	59.5	11.2	2442	1 CBP_HUMAN	Q92793 homo sapien
28	59	11.1	283	1 YJ81_YEAST	P47148 saccharomyc
29	59	11.1	460	1 5H2C_RAT	P08909 rattus norv
30	59	11.1	619	1 NBL4_BRARE	O57457 brachydanio
31	59	11.1	654	1 ICAL_RAT	P27321 rattus norv
32	58.5	11.0	351	1 SRYB_DROME	P07665 drosophila
33	58.5	11.0	698	1 TRFE_HUMAN	P02787 homo sapien

ALIGNMENTS

RESULT 1				
GP72_MOUSE				
ID	GP72_MOUSE	STANDARD;	PRT;	423 AA.
AC	P30731;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Probable G protein-coupled receptor GPR72 precursor (Glucocorticoid-induced receptor).			
GN	GPR72 OR GIR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92123228; PubMed=1663214;			
RA	Harrigan M.T., Campbell N.F., Bourgeois S.;			
RT	"Identification of a gene induced by glucocorticoids in murine T-cells: a potential G protein-coupled receptor.";			
RL	Mol. Endocrinol. 5:1331-1338(1991).			
CC	-!- FUNCTION: ORPHAN RECEPTOR. COULD BE A NEUROPEPTIDE Y RECEPTOR.			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-!- ALTERNATIVE PRODUCTS: MULTIPLE FORMS OF GIR MAY ARISE BY ALTERNATIVE SPLICING.			
CC	-!- TISSUE SPECIFICITY: BRAIN, THYMUS AND SPLEEN.			
CC	-!- INDUCTION: BY GLUCOCORTICOIDS AND CAMP; IN T-CELLS.			
CC	-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M80481; AAA17882.1; --			
DR	MGD; MGI:95712; Gir.			
DR	InterPro; IPR000276; GPCR_Rhodpsn.			
DR	Pfam; PF00001; 7tm_1; 1.			
DR	PRINTS; PR00237; GPCRRHODOPSIN.			
DR	PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.			
DR	PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.			
KW	G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;			
KW	Alternative splicing.			
FT	SIGNAL	1	17	POTENTIAL.
FT	CHAIN	18	423	PROBABLE G PROTEIN-COUPLED RECEPTOR
FT				GPR72.
FT	DOMAIN	18	71	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	72	92	1 (POTENTIAL).
FT	DOMAIN	93	107	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	108	129	2 (POTENTIAL).
FT	DOMAIN	130	145	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	146	167	3 (POTENTIAL).
FT	DOMAIN	168	186	CYTOPLASMIC (POTENTIAL).





```

[2]
RN  SEQUENCE FROM N.A.
RP  STRAIN=K12;
RC  MEDLINE=97251357; PubMed=9097039;
RX  Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA  Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA  Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA  Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,
RA  Oshima T., Saito N., Sampei G., Seki Y., Sivasundaram S.,
RA  Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA  Yamamoto Y., Horiuchi T.;
RT  "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RL  corresponding to the 28.0-40.1 min region on the linkage map.";
RN  DNA Res. 3:363-377(1996).

[3]
RN  SEQUENCE FROM N.A.
RP  STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC  MEDLINE=21074935; PubMed=11206551;
RX  Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

```

RESULT 4			
C941_VICSA	STANDARD;	PRT;	514 AA.
ID C941_VICSA			
AC O81117;			
DT 15-JUN-2002 (Rel. 41, Created)			
DT 15-JUN-2002 (Rel. 41, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DE Cytochrome P450 94A1 (EC 1.14.-.-) (P450-dependent fatty acid			
DE omega-hydroxylase).			
GN CYP94A1 OR VAGH11.			
OS <i>Vicia sativa</i> (Spring vetch) (Tare).			
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; <i>Vicia</i> .			
OX NCBI_TaxID=3908;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Seedling;			
RX MEDLINE=98264856; PubMed=9601090;			
RA Tijet N., Helvig C., Pinot F., Le Bouquin R., Lesot A., Durst F.,			
RA Salauen J.-P., Benveniste I.;			

RT "Functional expression in yeast and characterization of a clofibrate-  
RT inducible plant cytochrome P-450 (CYP94A1) involved in cutin monomers  
RT synthesis.";  
RL Biochem. J. 332:583-589(1998).  
CC -!- FUNCTION: Catalyzes the omega-hydroxylation of various fatty acids  
CC (FA) from 10 to 18 carbon atoms. The substrate specificity is  
CC higher for laurate > palmitate > myristate > linolenate >  
CC linoleate > oleate > caprate. May play a minor role in cutin  
CC synthesis and could be involved in plant defense.  
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By  
CC similarity).  
CC -!- INDUCTION: By clofibrate.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF030260; AAD10204.1; -.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; p450; 1.  
DR PRINTS; PR00385; P450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; FALSE\_NEG.  
KW Oxidoreductase; Monooxygenase; Heme; Transmembrane;  
KW Endoplasmic reticulum; Multigene family.  
FT TRANSMEM 7 29 POTENTIAL.  
FT BINDING 458 458 HEME (BY SIMILARITY).  
SQ SEQUENCE 514 AA; 59110 MW; 3D9361380D6C3B0E CRC64;

Query Match 12.2%; Score 65; DB 1; Length 514;  
Best Local Similarity 20.0%; Pred. No. 10;  
Matches 34; Conservative 21; Mismatches 33; Indels 82; Gaps 9;

QY 6 RDHASLGDSSETLSQ--TELK---KERRKKRE-----RKFK 36  
Db 246 KKYFNIGSEKRLKEAVTEVRSFAKKLVREKKRELEKSSLETEDMLSRFLSSGHSDDFV 305  
QY 37 ANCGIDFI-----IFWIFWIL-----LFSSH 57  
Db 306 ADIVISFILAGKDTTSAALTWFFWLLWKNPRVEEIVNELSKKSELMVYDEVKEMVYTHA 365  
QY 58 WIQESL-LCPSPSPKEVTCREMLTGGCLP--WATRS-----HLGRRK 95  
Db 366 ALSESMRLYPVPMD--SKEAVNDVLPDGWVVKKGTTIVYHYVYAMGRMK 413

RESULT 5  
APJ\_XENLA  
ID APJ\_XENLA STANDARD; PRT; 353 AA.  
AC P79960; P70058;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE G protein-coupled receptor APJ homolog (Angiotensin receptor related  
DE protein) (Mesenchyme-associated serpentine receptor).  
GN X-MSR OR XAngiol.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97109511; PubMed=8951791;  
RA Devic E., Paquereau L., Vernier P., Knibiehler B., Audigier Y.;  
RT "Expression of a new G protein-coupled receptor X-msr is associated  
RL with an endothelial lineage in Xenopus laevis.";  
RN Mech. Dev. 59:129-140(1996).  
[2]

RP SEQUENCE OF 1-303 FROM N.A.  
RA Saha M.S., Oakes J.A., Miles R.R.;  
RT "XAngiol, a novel Xenopus gene, is expressed in vascular precursor  
RT cells.";  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Putative receptor for an apelin-like peptide coupled to  
CC G proteins that inhibit adenylate cyclase activity.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: At the gastrula stage, exclusively expressed  
CC in the mesodermal layer and at the neurula stage in the lateral  
CC plate mesoderm. Larval expression is observed in the endothelium  
CC of the primary blood vessels and the forming heart.  
CC -!- DEVELOPMENTAL STAGE: First expressed at the late blastula stage,  
CC increases during gastrulation and remains constant between neurula  
CC and larva stages.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X93045; CAA63612.1; -.  
DR EMBL; U72029; AAB17004.1; -.  
DR InterPro; IPR000276; GPCR\_Rhodpsn.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PRINTS; PR00237; GPCRRHODOPSN.  
DR PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; 1.  
DR PROSITE; PS00262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
KW G-protein coupled receptor; Glycoprotein; Transmembrane.  
FT DOMAIN 1 38 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 39 61 1 (POTENTIAL).  
FT DOMAIN 62 72 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 73 95 2 (POTENTIAL).  
FT DOMAIN 96 109 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 110 131 3 (POTENTIAL).  
FT DOMAIN 132 150 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 151 173 4 (POTENTIAL).  
FT DOMAIN 174 206 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 207 229 5 (POTENTIAL).  
FT DOMAIN 230 248 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 249 270 6 (POTENTIAL).  
FT DOMAIN 271 294 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 295 317 7 (POTENTIAL).  
FT DOMAIN 318 353 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 19 19 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 150 150 P -> S (IN REF. 2).  
SQ SEQUENCE 353 AA; 40303 MW; 1BF757D865057621 CRC64;

Query Match 11.8%; Score 63; DB 1; Length 353;  
Best Local Similarity 30.6%; Pred. No. 11;  
Matches 15; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 11 LGDSETLSQTELKRRKKRERKFOANCIDFIIFWIFWILLFSHHWI 59  
Db 228 IGGKVTMHFQNLKKEQKKRLKIITLVVFAICWLPFHILKTIHFL 276

## RESULT 6

ATPD\_VIBAL  
ID ATPD\_VIBAL STANDARD; PRT; 177 AA.  
AC P12987;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE ATP synthase delta chain (EC 3.6.3.14).  
GN ATPH OR UNCH.  
OS Vibrio alginolyticus.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.





FT	REPEAT	449	476	1-2.
FT	REPEAT	477	504	1-3.
FT	REPEAT	505	532	1-4.
FT	REPEAT	533	560	1-5.
FT	REPEAT	561	588	1-6.
FT	REPEAT	589	616	1-7.
FT	REPEAT	617	644	1-8.
FT	REPEAT	645	678	1-9.
FT	REPEAT	679	706	1-10.
FT	REPEAT	707	742	1-11.
FT	REPEAT	743	771	1-12.
FT	REPEAT	772	796	1-13.
FT	REPEAT	797	832	1-14.
FT	DOMAIN	938	1507	23 X 23 AA APPROXIMATE TANDEM REPEATS
FT	REPEAT	938	961	2-1.
FT	REPEAT	962	985	2-2.
FT	REPEAT	986	1021	2-3.
FT	REPEAT	1022	1044	2-4.
FT	REPEAT	1045	1067	2-5.
FT	REPEAT	1068	1090	2-6.
FT	REPEAT	1091	1121	2-7.
FT	REPEAT	1122	1144	2-8.
FT	REPEAT	1145	1167	2-9.
FT	REPEAT	1168	1197	2-10.
FT	REPEAT	1198	1227	2-11.
FT	REPEAT	1228	1250	2-12.
FT	REPEAT	1251	1273	2-13.
FT	REPEAT	1274	1296	2-14.
FT	REPEAT	1297	1319	2-15.
FT	REPEAT	1320	1342	2-16.
FT	REPEAT	1343	1368	2-17.
FT	REPEAT	1369	1391	2-18.
FT	REPEAT	1392	1416	2-19.
FT	REPEAT	1417	1439	2-20.
FT	REPEAT	1440	1461	2-21.
FT	REPEAT	1462	1484	2-22.
FT	REPEAT	1485	1507	2-23.
FT	VARIANT	1145	1197	MISSING (IN SHORT FORM).
FT	VARIANT	1251	1273	MISSING (IN SHORT FORM).
FT	CONFLICT	1399	1399	E -> G (IN REF. 2).
SQ	SEQUENCE	1549 AA;	201173 MW;	E72FB9FF1326E54E CRC64;

Query Match 11.7%; Score 62; DB 1; Length 1549;

Best Local Similarity 37.1%; Pred. No. 69;

Matches 13; Conservative 9; Mismatches 13; Indels 0; Gap

QY	1	EVEVSRDHASLGDSETLSQLTKKKRKKRKF	35
Db	418	EVELQREERLQREEEQLQREERKKRRQREKQY	452

RESULT 9

ACD1\_CLOPE

ID ACD1\_CLOPE STANDARD; PRT; 218 AA.

AC Q8XMY2;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative acyl carrier protein phosphodiesterase 1 (EC 3.1.4.14) (ACP phosphodiesterase 1).

GN CPE0556.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI\_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=13 / Type A;

RX PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;

RT "Complete genome sequence of Clostridium perfringens, an anaerobic

RT flesh-eater.";



Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

-!- FUNCTION: Converts holo-ACP to apo-ACP by hydrolytic cleavage of the phosphopantetheine prosthetic group from ACP (By similarity).

-!- CATALYTIC ACTIVITY: Holo-[acyl-carrier protein] + H(2)O = 4'-phosphopantetheine + apo-[acyl-carrier protein].

-!- SIMILARITY: BELONGS TO THE ACPD FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AP003187; BAB80262.1; -

Hypothetical protein; Hydrolase; Complete proteome.

SEQUENCE 218 AA; 24965 MW; 82E54E1A03EE253F CRC64;

Query Match 11.6%; Score 61.5; DB 1; Length 218;

Best Local Similarity 24.8%; Pred. No. 10;

Matches 29; Conservative 17; Mismatches 42; Indels 29; Gaps 4;

Qy 1 EVEVSRD-----HASLGDSETLQSOTELRKKRKKRKFQANCIDFIIFWI 48

Db 41 EVDLYKDFIPRLKHFSGRSSVVDCTASIDAKTKEVKRIIELSEQFNADLYVIAAP 100

Qy 49 FWILFHHW-----IQESLLCPPSPKVEV-----TCREML---TGGCLPWATR 88

Db 101 LWTLSFPAPLKEYIDCVQNKITIKVTPEEVKGLDDKQREMYVIQSSGGEIPWLTK 157

RESULT 10

SH7\_HUMAN STANDARD; PRT; 479 AA.

AC P34969; P78516; P78336; P78372;

DT 01-FEB-1994 (Rel. 28, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 5-hydroxytryptamine 7 receptor (5-HT-7) (5-HT-X) (Serotonin receptor) (5HT7).

GN HTR7.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.

RX MEDLINE=97238071; PubMed=9084407;

RA Heidmann D.E.A., Metcalf M.A., Kohen R., Hamblin M.W.;

RT "Four 5-hydroxytryptamine7 (5-HT7) receptor isoforms in human and rat produced by alternative splicing: species differences due to altered intron-exon organization."

RT J. Neurochem. 68:1372-1381(1997).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM A).

RC TISSUE=Placenta, and Fetal brain;

RX MEDLINE=94043137; PubMed=8226867;

RA Bard J.A., Zgombick J.M., Adham N., Vaysse P., Brancheck T.A., Weinshank R.L.;

RT "Cloning of a novel human serotonin receptor (5-HT7) positively linked to adenylate cyclase."

RT J. Biol. Chem. 268:23422-23426(1993).

RN [3]

RP VARIANTS LYS-92 AND LEU-279.

RX MEDLINE=97298817; PubMed=9154233;

RA Erdmann J., Nothen M.M., Shimron-Ababanell D., Rietschel M., Albus M., Borrmann M., Maier W., Franzek E., Korner J., Weigelt B., Fimmers R., Propping P.;

RT "The human serotonin 7 (5-HT7) receptor gene: genomic organization and systematic mutation screening in schizophrenia and bipolar affective disorder."

RL Mol. Psych. 1:392-397(1996).

-!- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR 5-HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTIONS AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT STIMULATE ADENYLATE CYCLASE.

-!- SUBCELLULAR LOCATION: Integral membrane protein.

-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; D (SHOWN HERE), A AND B; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORM D IS THE LONGEST ISOFORM BUT ISOFORMS A AND B APPEAR TO BE EXPRESSED AT HIGHER LEVELS.

-!- TISSUE SPECIFICITY: ISOFORM A IS THE PREDOMINANT ISOFORM IN SPLEEN, CAUDATE AND HIPPOCAMPUS. ISOFORM B IS EXPRESSED AT LOWER LEVELS, AND ISOFORM D IS A MINOR ISOFORM.

-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U68487; AAB48393.1; -

EMBL; U68488; AAB48394.1; -

EMBL; U68492; AAF07218.1; -

EMBL; U68493; AAF07217.1; -

EMBL; U68493; AAF07217.1; JOINED.

EMBL; U68493; AAB48397.2; -

EMBL; U68492; AAB48397.2; JOINED.

EMBL; L21195; AAC37538.1; -

Genew; HGNC:5302; HTR7.

MIM; 182137; -

InterPro; IPR000276; GPCR\_Rhodpsn.

Pfam; PF00001; 7tm\_1; 1.

PRINTS; PR00237; GPCRHDOPSN.

PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; 1.

PROSITE; PS00262; G\_PROTEIN\_RECEP\_F1\_2; 1.

KW G-protein coupled receptor; Transmembrane; Glycoprotein; Multigene family; Lipoprotein; Palmitate; Alternative splicing; Polymorphism.

FT DOMAIN 1 83 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 84 104 1 POTENTIAL.

FT DOMAIN 105 117 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 118 138 2 POTENTIAL.

FT DOMAIN 139 157 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 158 178 3 POTENTIAL.

FT DOMAIN 179 201 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 202 222 4 POTENTIAL.

FT DOMAIN 223 236 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 237 257 5 POTENTIAL.

FT DOMAIN 258 325 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 326 346 6 POTENTIAL.

FT DOMAIN 347 367 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 368 388 7 POTENTIAL.

FT DOMAIN 389 479 CYTOPLASMIC (POTENTIAL).

FT CARBOHYD 5 5 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 66 66 N-LINKED (GLCNAC. .) (POTENTIAL).

FT DISULFID 155 231 BY SIMILARITY.

FT LIPID 401 401 PALMITATE (POTENTIAL).

FT VARSPLIC 433 479 MISSING (IN ISOFORM B).

FT VARSPLIC 433 479 RACTRRVLLRPEKRPVSVVQLQSPDHHNLADKMLTVEK

FT VARIANT 92 92 KVMTHD -> QNADYCRKKGHDS (IN ISOFORM A).

FT VARIANT 279 279 T -> K.

FT VARIANT 279 279 /FTID=VAR\_012995.

FT VARIANT 279 279 P -> L.

FT VARIANT 279 279 /FTID=VAR\_012996.

SQ SEQUENCE 479 AA; 53554 MW; 1F62E985EAE1F23 CRC64;

Query Match 11.6%; Score 61.5; DB 1; Length 479;

Best Local Similarity 29.4%; Pred. No. 23;

Matches 25; Conservative 8; Mismatches 29; Indels 23; Gaps 5;

Qy 15 ETLSQLTKRKKK----KRERKFQANGCIDFIIFWIFWILLFHHWQLQSPSPK 70

Db 301 ECANLSRLKHERKNISIFKREQKAATLGIIVGFTVCWLPFF-----LLS----- 347

QY 71 EVTCREMLTG---GCLP-WATRSHL 91

Db 348 --TARPFICGTSCSCIPLWVERTFL 370

RESULT 11

RS1\_RHIME STANDARD; PRT; 568 AA.

AC P14129;

DT 01-JAN-1990 (Rel. 13, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 30S ribosomal protein S1.

GN RPSA OR R00255 OR SMC00335.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI\_TaxID=382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RCR2011 / SU47;

RX MEDLINE=88217521; PubMed=3368316;

RA Schnier J., Thamm S., Lurz R., Hussain A., Faist G.,

RA Dobrinski B.;

RT "Cloning and characterization of a gene from Rhizobium meliloti 2011

RT coding for ribosomal protein S1.";

RL Nucleic Acids Res. 16:3075-3089(1988).

[2]

RN SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Guzy J., Bothe G., Ampe F., Batut J.,

RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,

RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,

RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;

RT "Analysis of the chromosome sequence of the legume symbiont

RT Sinorhizobium meliloti strain 1021.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).

CC -!- FUNCTION: BINDS MRNA; THUS FACILITATING RECOGNITION OF THE

CC INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT

CC SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE S1P FAMILY OF RIBOSOMAL PROTEINS.

CC -!- SIMILARITY: CONTAINS 6 S1 MOTIF DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X07528; CAA30404.1; -.

CC EMBL; AL591783; CAC41692.1; -.

CC PIR; S01055; R3ZR1.

CC HSSP; P05055; 1SRO.

CC InterPro; IPR00110; Ribosomal\_S1.

CC InterPro; IPR003029; S1.

CC Pfam; PF00575; S1; 6.

CC PRINTS; PRO0681; RIBOSOMALS1.

CC SMART; SM00316; S1; 6.

CC TIGRFAMS; TIGR00717; rpsA; 1.

CC PROSITE; PS50126; S1; 6.

KW Ribosomal protein; Repeat; RNA-binding; Complete proteome.

FT DOMAIN 27 93

FT S1 MOTIF 1.

FT DOMAIN 111 177

FT S1 MOTIF 2.

FT DOMAIN 198 266

FT S1 MOTIF 3.

FT DOMAIN 283 353

FT S1 MOTIF 4.

FT DOMAIN 370 440

FT S1 MOTIF 5.

FT DOMAIN 459 530 S1 MOTIF 6.

FT CONFLICT 88 A -> R (IN REF. 1).

FT CONFLICT 153 LMHNPPQFEI -> ADAQPAALRN (IN REF. 1).

FT CONFLICT 235 N -> K (IN REF. 1).

FT CONFLICT 241 L -> Q (IN REF. 1).

FT CONFLICT 551 A -> R (IN REF. 1).

SQ SEQUENCE 568 AA; 62640 MW; AB0858204273A7B8 CRC64;

Query Match 11.6%; Score 61.5; DB 1; Length 568;

Best Local Similarity 37.0%; Pred. No. 27;

Matches 17; Conservative 12; Mismatches 16; Indels 1; Gaps 1;

QY 1 EVEVSRDHASLGDSETLSQTELRKKKKRKRKFKQANCGIDFIIF 46

Db 75 EVYVERIENALGEA-VLSREKARREESWQRLEVKFEAGERVEGIIF 119

RESULT 12

POLG\_HRV89

ID POLG\_HRV89 STANDARD; PRT; 2164 AA.

AC P07210; Q82096; Q82097; Q82098; Q82099; Q82100; Q82101; Q82102;

AC Q82103; Q82104; Q82105;

DT 01-APR-1988 (Rel. 07, Created)

DT 01-APR-1988 (Rel. 07, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins

DE P2A TO P2C, P3A; Genome-linked protein VPg; Picornain 3C

DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D

DE (EC 2.7.7.48)].

OS Human rhinovirus 89 (HRV-89).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Rhinovirus.

OX NCBI\_TaxID=12132;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87204179; PubMed=3033653;

RA Duechler M., Skern T., Sommergruber W., Neubaue C., Gruendler P.,

RA Fogy I., Blaas D., Kuechler E.;

RT "Evolutionary relationships within the human rhinovirus genus:

RT comparison of serotypes 89, 2, and 14.";

RL Proc. Natl. Acad. Sci. U.S.A. 84:2605-2609(1987).

CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN

CC Q/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.

CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the

CC poliovirus polyprotein. In other picornavirus reactions Glu may be

CC substituted for Gln, and Ser or Thr for Gly.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +

CC {RNA}(N).

CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,

CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,

CC VP3, AND VP4.

CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M16248; AAA45762.1; -.

CC EMBL; A10937; CAA00931.1; -.

CC PIR; A29862; GNNY89.

CC MEROPS; C03.007; -.

CC MEROPS; C03.021; -.

CC InterPro; IPR000199; Cys-protease-3C.

CC InterPro; IPR003138; Pico\_P1A.

CC InterPro; IPR000081; Pico\_P2A.

CC InterPro; IPR002527; Pico\_P2B.

CC InterPro; IPR000605; RNA\_helicase.

CC InterPro; IPR001205; RNA\_pol\_P3D.













```

KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT NON_TER 396 396
SQ SEQUENCE 396 AA; 44050 MW; 1F7EC53DB27A8669 CRC64;

Query Match
Best Local Similarity 20.2%; Score 72; DB 15; Length 396;
Matches 25; Conservative 19; Mismatches 32; Indels 48; Gaps 4;

QY 3 EVSRDHASLGDSSE---TLSTQLRKKKKRKKRKKFQA-----NCGIDFIIF 46
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 DIRQAHCSLGRAEWNNTLKQIVIKLKEQFNKTIIVFKQSSGGDPEIVMHSFNCGGEF--F 259
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 47 WIFWILLFSSH-----IQESLLCPPSPKVEVTCRE 76
   : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 260 YCNTTQLFNSTWNVTGLNTEGDDTIILPCRIKQIINMWQEVGKAMYAPPISGQIRCSS 319
   : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 MLTG 80
   : | |
Db 320 NITG 323

RESULT 6
O11505
ID O11505 PRELIMINARY; PRT; 396 AA.
AC O11505;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Balotta C., Riva C., Violin M., Colombo C., Galli M., Moroni M.;
RT "Analysis of the V1-V5 env region in Long-Term Non Progressor and in a
RT rapid progressor infected individuals."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U95408; AAB52754.1; -
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT NON_TER 396 396
SQ SEQUENCE 396 AA; 44120 MW; EF6AD3F0B027FB21 CRC64;

Query Match
Best Local Similarity 20.2%; Score 72; DB 15; Length 396;
Matches 25; Conservative 19; Mismatches 32; Indels 48; Gaps 4;

QY 3 EVSRDHASLGDSSE---TLSTQLRKKKKRKKRKKFQA-----NCGIDFIIF 46
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 DIRQAHCSLGRAEWNNTLKQIVIKLKEQFNKTIIVFKQSSGGDPEIVMHSFNCGGEF--F 259
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 47 WIFWILLFSSH-----IQESLLCPPSPKVEVTCRE 76
   : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 260 YCNTTQLFNSTWNVTGLNTEGDDTIILPCRIKQIINMWQEVGKAMYAPPISGQIRCSS 319
   : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 MLTG 80
   : | |
Db 320 NITG 323

RESULT 7
Q9U0G5
ID Q9U0G5 PRELIMINARY; PRT; 2277 AA.
AC Q9U0G5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Var, MAL4P2.58.
```

```

GN MAL4P2.58, VAR.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,
RA Quail M., Barrell B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035475; CAB62899.1; -
DR InterPro; IPR004258; PFEMP.
DR Pfam; PF03011; PFEMP; 2.
SQ SEQUENCE 2277 AA; 256122 MW; E4262CCA69DDEF93 CRC64;

Query Match
Best Local Similarity 23.9%; Score 71; DB 5; Length 2277;
Matches 21; Conservative 15; Mismatches 32; Indels 20; Gaps 5;

QY 22 LRKK--ERKKRERKFOACGIDFIIFWIFWILLFHHWQESLL-CPPSPKVEVTCREML 78
   | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 657 LQKNGEENGKKQKTYN-----DFNFVVAHMLKDSIHWRTKKIKGCLKNGKAICTDKC 711
   | : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 79 TGGCL-----PWAT-RSHLGR 94
   | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 712 KGCKCFERWVEQKREWTNIKEHFGKQ 739
   | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
O16652
ID O16652 PRELIMINARY; PRT; 684 AA.
AC O16652;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE T06D4.4 protein.
GN T06D4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Le T.T., Goela D.;
RT "The sequence of C. elegans cosmid T06D4."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016673; AAB66123.1; -
DR InterPro; IPR000718; Peptidase_M13.
DR Pfam; PF01431; Peptidase_M13; 1.
SQ SEQUENCE 684 AA; 79176 MW; C2A8C605CB007A08 CRC64;
```





```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; Z68296; CAA92591.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01574; TUBBYPROTEIN.
DR ProDom; PD000001; Euk_pkinase; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1851 AA; 199918 MW; 71BD78818EF6A5A5 CRC64;

Query Match      12.6%; Score 67; DB 5; Length 1851;
Best Local Similarity 33.3%; Pred. No. 38;
Matches 14; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

QY      2 VEVSRDHASLGDSSETLSQTELKRRKKRERKFKQANCIDF 43
Db      617 VEIKNRDADLNDLNVETQMLRVYDEKKRQYRFKNEGLQF 658

```

Search completed: July 9, 2003, 15:06:30  
Job time : 46.7788 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 8.92035 Seconds  
(without alignments)  
239.005 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FOANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	100.0	16	AAE13851	Human T cell epitope
2	50	51.0	504	AA1980	Enzyme EPS11 invol
3	50	51.0	504	AA1981	Amino acid sequenc
4	47	48.0	207	AAW99067	Streptococcus pneu
5	47	48.0	326	AAW99066	Streptococcus pneu
6	47	48.0	326	AAW01092	CFE 95 protein seq
7	46	46.9	50	AAW84080	Human immune/haema
8	46	46.9	64	AAW57264	Human prostate can
9	46	46.9	123	AAW54009	Anti-CD4 antibody
10	45	45.9	285	AAW51648	Rat GPCR polypepti

11	45	45.9	298	15	AAE13851	Human thoracic aor
12	45	45.9	298	17	AAW02727	Human thoracic aor
13	45	45.9	323	23	AAU76411	Novel human G-prot
14	45	45.9	327	23	AAE14729	Human G-protein co
15	45	45.9	343	17	AAE14729	Human G-protein co
16	45	45.9	343	17	AAE14729	Human G-protein co
17	45	45.9	343	17	AAE14729	Human G-protein co
18	45	45.9	343	22	AAE14729	Human membrane or
19	45	45.9	343	23	AAE14729	RTA-like G protein
20	45	45.9	1299	22	AAE14729	Drosophila melanog
21	45	45.9	1299	22	AAE14729	Human polypeptide
22	45	45.9	1299	22	AAE14729	Human polypeptide
23	45	45.9	1299	22	AAE14729	Drosophila melanog
24	45	45.9	1299	22	AAE14729	Human thyroid stim
25	45	45.9	1299	22	AAE14729	Humanised antibody
26	45	45.9	1299	22	AAE14729	Human gene 43 enco
27	45	45.9	1299	22	AAE14729	Human polypeptide
28	45	45.9	1299	22	AAE14729	Human stromal cell
29	45	45.9	1299	22	AAE14729	Human stomach canc
30	45	45.9	1299	22	AAE14729	Human secreted pro
31	45	45.9	1299	22	AAE14729	Human receptor mol
32	45	45.9	1299	22	AAE14729	Human gene 43 enco
33	45	45.9	1299	22	AAE14729	Human stromal cell
34	45	45.9	1299	22	AAE14729	Human secreted pro
35	45	45.9	1299	22	AAE14729	Human colon cancer
36	45	45.9	1299	22	AAE14729	Human polypeptide
37	45	45.9	1299	22	AAE14729	Helicobacter pylor
38	45	45.9	1299	22	AAE14729	H. pylori GHPO 534
39	45	45.9	1299	22	AAE14729	Helicobacter pylor
40	45	45.9	1299	22	AAE14729	Herbicidally activ
41	45	45.9	1299	22	AAE14729	Human gene 50 enco
42	45	45.9	1299	22	AAE14729	Human albumin fusi
43	45	45.9	1299	22	AAE14729	C6 human sfv antib
44	45	45.9	1299	22	AAE14729	C6 human sfv antib
45	45	45.9	1299	22	AAE14729	C6 human sfv antib

ALIGNMENTS

RESULT 1  
AAE13851  
ID AAE13851 standard; peptide; 16 AA.  
XX  
AC AAE13851;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Human T cell epitope related to lung tumour-specific protein.

Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
antisense-therapy; vaccine; immune response; lung cancer;  
T cell epitope.

XX Homo sapiens.

XX WO200172295-A2.

XX 04-OCT-2001.

XX 28-MAR-2001; 2001WO-US09991.

XX 29-MAR-2000; 2000US-0538037.

XX 05-JUN-2000; 2000US-0588937.

XX 18-AUG-2000; 2000US-0640878.

XX 22-SEP-2000; 2000US-234517P.

XX 01-NOV-2000; 2000US-0704512.

XX 14-DEC-2000; 2000US-0738973.

XX (CORI-) CORIXA CORP.

PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;

XX WPI; 2001-639201/73.

XX New human lung-specific polynucleotides and polypeptides for the

PT diagnosis and treatment of disease e.g. lung cancer -

PT

XX Claim 2; Page 378; 378pp; English.

XX

CC The invention relates to isolated lung tumour-specific proteins and

CC their corresponding cDNA molecules. Lung tumour-specific proteins and

CC their antigen-presenting cells are useful for stimulating and/or

CC expanding T cells specific for a tumour protein, and for inhibiting

CC the development of cancer. The invention also relates to a composition

CC useful for stimulating an immune response, and for treating cancer. The

CC lung tumour specific oligonucleotide is useful in gene therapy and for

CC diagnosis, detection and treatment of lung cancer. The present sequence

CC is human T cell epitope related to lung tumour-specific protein.

XX

SQ Sequence 16 AA;

Query Match 100.0%; Score 98; DB 22; Length 16;

Best Local Similarity 100.0%; Pred. No. 1.1e-08;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16

Db 1 FQANCGIDFIIFWIFW 16

RESULT 2

AAAY54085

ID AAY54085 standard; Protein; 504 AA.

XX

AC AAY54085;

XX

DT 27-MAR-2000 (first entry)

XX

DE Enzyme EPS11 involved in exopolysaccharide biosynthesis.

XX

KW Exopolysaccharide; EPS; ESP enzyme; EPS1; EPS2; EPS3; EPS4; EPS5; EPS6;

KW EPS7; EPS8; EPS9; EPS10; EPS11; Lactobacillus helveticus strain LH59;

KW activated D-galactose pyranose; saccharide; beta-glycosyltransferase;

KW undecaprenyl-phosphate-glycosyl-1-phosphate-transferase;

KW alpha-glycosyltransferase; EPS polymerase; glycosyltransferase;

KW phosphofuranose; transporter; food; fermented milk product; yoghurt;

KW cheese; flavour stability; organoleptic property.

XX

OS Lactobacillus helveticus.

XX

PN WO9962316-A2.

XX

PD 09-DEC-1999.

XX

PF 22-APR-1999; 99WO-EP02841.

XX

PR 22-APR-1998; 98EP-0201310.

PR 22-APR-1998; 98EP-0201311.

PR 22-APR-1998; 98EP-0201312.

XX

PA (NEST ) SOC PROD NESTLE SA.

XX

PI Stingele F, Germond JE, Lamothe G;

XX

DR WPI; 2000-097267/08.

DR N-PSDB; AAZ45259, AAY54082, AAY54083, AAY54084; AAY54085.

XX

XX New recombinant enzymes for synthesis of exopolysaccharides,

PT particularly in lactic acid bacteria, for improving properties of

PT fermented milk products -

XX

XX Example 8; Page 130-131; 162pp; French.

PS

XX AAY54075-85 represent enzymes involved in the biosynthesis of

CC

CC exopolysaccharides (EPS). These enzymes are designated EPS1-EPS11. and

CC are encoded by open reading frames eps1-eps11. The enzymes are isolated

CC from Lactobacillus helveticus strain LH59. The proteins are used

CC in a method for the synthesis of EPS, which includes at least one step

CC of forming a bond (alpha or beta-isomer) between C-1 (carrying the

CC reducing aldehyde function, of an activated D-galactose pyranose), and

CC a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis

CC of EPS occurs with, in each step, addition of a new sugar unit, through

CC its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar

CC unit, present at the end of a chain of sugar residues bonded to the

CC primer. EPS1 has homology with an undecaprenyl-phosphate-glycosyl-1-

CC phosphate-transferase; EPS2, EPS3 and EPS4 are alpha-

CC glycosyltransferases; EPS5 and EPS6 have homology with a beta-

CC glycosyltransferase; EPS7 has homology with an EPS polymerase; EPS8

CC is a glycosyltransferase; EPS9 catalyses the transfer of phosphofuranose

CC onto the following repetitive unit; EPS10 transports the polysaccharides

CC formed; and EPS11 is involved in the synthesis and export of formed

CC polysaccharides. The EPS enzyme are used to improve properties of foods,

CC particularly fermented milk products such as yoghurt and cheese,

CC e.g. their organoleptic properties and flavour stability.

XX

SQ Sequence 504 AA;

Query Match 51.0%; Score 50; DB 21; Length 504;

Best Local Similarity 37.5%; Pred. No. 8.6;

Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16

Db 114 YRSKCIVTFMHFWFFW 129

RESULT 3

AAAY43787

ID AAY43787 standard; Protein; 504 AA.

XX

AC AAY43787;

XX

DT 11-FEB-2000 (first entry)

XX

DE Amino acid sequence of the eps11 of Lactobacillus helveticus LH59.

XX

KW eps operon; Lactobacillus helveticus LH59; enzyme; eps1; eps2;

KW eps3; eps4; eps5; eps6; eps7; eps8; eps9; eps10; eps11;

KW exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;

KW probiotic; foodstuff; organoleptic quality; flavour;

KW lactic acid bacteria; acidified milk product; yoghurt; cheese.

XX

OS Lactobacillus helveticus.

XX

PN WO9954475-A2.

XX

PD 28-OCT-1999.

XX

PF 22-APR-1999; 99WO-EP03011.

XX

PR 22-APR-1998; 98EP-0201310.

PR 22-APR-1998; 98EP-0201311.

PR 22-APR-1998; 98EP-0201312.

XX

PA (NEST ) SOC PROD NESTLE SA.

XX

PI Stingele F, Germond JE, Lamothe G;

XX

DR WPI; 2000-013255/01.

DR N-PSDB; AAZ30356, AAY43784, AAY43785, AAY43786, AAY43787.

XX

XX New recombinant enzymes for biosynthesis of exopolysaccharides having

PT e.g. antitumor or probiotic properties or useful in fermented milk

PT products -

XX

XX Example 8; Page 131-132; 163pp; French.

PS

XX

CC AAY43777-87 represent enzymes encoded by the eps operon of Lactobacillus  
CC helveticus LH59. The operon contains 11 open reading frames, and  
CC encodes enzymes (eps1, eps2, eps3, eps4, eps5, eps6, eps7, eps8, eps9  
CC and eps11) that are involved in the biosynthesis of exopolysaccharides  
CC (EPS). The enzymes catalyse the formation of specific intersugar bonds.  
CC The enzymes catalyse a process that includes at least one step of  
CC forming a bond (in alpha or beta anomeric form) between C1, carrying  
CC the reducing aldehyde group of an activated D-Galp (galactose in  
CC pyranose form), and a phosphate group on a lipophilic or proteinaceous  
CC primer. The enzymes are used to produce EPS that have antitumor or  
CC probiotic properties or are used in foodstuffs to improve organoleptic  
CC qualities and flavour. When expressed by lactic acid bacteria, EPS  
CC impart a free-flowing character and/or a smooth, creamy texture to  
CC acidified milk products (yoghurt or cheese).

XX Sequence 504 AA;

SQ Query Match 51.0%; Score 50; DB 21; Length 504;  
Best Local Similarity 37.5%; Pred. No. 8.6;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIIFWIFW 16  
Db 114 YRSKCIVTFMHFFW 129  
::: | : | : | | |

RESULT 4

AAW99067  
ID AAW99067 standard; Protein; 207 AA.

AC AAW99067;

XX 14-MAY-1999 (first entry)

DT Streptococcus pneumoniae mray transferase ORF protein sequence.

DE Streptococcus pneumoniae; mray; transferase; infection; pneumonia;  
KW phospho-N-acetylmuramoyl-pentapeptide-transferase family; otitis media;  
KW bacteraemia; conjunctivitis; sinusitis; pleural empyema; endocarditis;  
KW meningitis; gene therapy.

XX Streptococcus pneumoniae.

OS EP897007-A2.

XX 17-FEB-1999.

PD 11-JUN-1998; 98EP-0304635.

XX 16-APR-1998; 98US-0061156.

PR 12-AUG-1997; 97US-0055467.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

XX Fueyo JL, Jaworski DD, Kosmatka AL, Lonetto MA;

PI Traini CM, Wang M;

XX WPI; 1999-134240/12.

DR N-PSDB; AAX18943.

XX New Phospho-N-acetylmuramoyl-pentapeptide-transferase (mray)

PT polypeptide and polynucleotide - useful as diagnostic reagents and

PT for prevention and treatment of Streptococcus pneumoniae

PT infections, which cause otitis media and meningitis

XX Claim 1; Page 4; 37pp; English.

PS The present sequence represents the ORF protein sequence of an isolated

CC Phospho-N-acetylmuramoyl-pentapeptide-transferase designated mray. Mray

CC patient sample. Mray PCR probes are useful for diagnosing diseases, and  
CC can characterise the response of the infectious organism to drugs. Mray  
CC proteins and polynucleotides are also useful for screening for  
CC antagonists, agonists and drugs against infectious micro-organisms. Mray  
CC agonists and antagonists are bacteriostatic and bacteriocidal compounds  
CC which can be used in treatment to enhance (agonist) or block (antagonist  
CC or antisense sequence) mray activity, therefore treating microbial  
CC diseases, especially Streptococcus pneumoniae diseases including otitis  
CC media, bacteraemia, conjunctivitis, pneumonia, sinusitis, pleural  
CC empyema, endocarditis and especially meningitis. Epitopes of mray  
CC proteins and polynucleotides are useful immunogens (vaccines) for  
CC producing anti-mray antibodies for prevention of bacterial infections,  
CC and mray polynucleotides can be used in genetic immunisation (gene  
CC therapy) to prevent infections. Mray proteins, polynucleotides and their  
CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and  
CC are useful for use on wounds and body implants to prevent bacterial  
CC infection.

XX Sequence 207 AA;

SQ Query Match 48.0%; Score 47; DB 20; Length 207;  
Best Local Similarity 43.8%; Pred. No. 11;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIIFWIFW 16

Db 24 YQVHLGIFYIVFALFW 39  
: | : | | : | : | |

RESULT 5

AAW99066

ID AAW99066 standard; Protein; 326 AA.

XX AAW99066;

XX 14-MAY-1999 (first entry)

DT Streptococcus pneumoniae mray transferase.

DE Streptococcus pneumoniae; mray; transferase; infection; pneumonia;  
KW phospho-N-acetylmuramoyl-pentapeptide-transferase family; otitis media;  
KW bacteraemia; conjunctivitis; sinusitis; pleural empyema; endocarditis;  
KW meningitis; gene therapy.

XX Streptococcus pneumoniae.

OS EP897007-A2.

XX 17-FEB-1999.

PD 11-JUN-1998; 98EP-0304635.

XX 16-APR-1998; 98US-0061156.

PR 12-AUG-1997; 97US-0055467.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

XX Fueyo JL, Jaworski DD, Kosmatka AL, Lonetto MA;

PI Traini CM, Wang M;

XX WPI; 1999-134240/12.

DR N-PSDB; AAX18942.

XX New Phospho-N-acetylmuramoyl-pentapeptide-transferase (mray)

PT polypeptide and polynucleotide - useful as diagnostic reagents and

PT for prevention and treatment of Streptococcus pneumoniae

PT infections, which cause otitis media and meningitis

XX Claim 1; Page 4; 37pp; English.

PS The present sequence represents an isolated Phospho-N-acetylmuramoyl-

CC pentapeptide-transferase designated mray. Mray polynucleotides and

CC proteins are useful for diagnosing susceptibility to diseases by  
CC detecting mutations or polymorphisms in the mray gene or analysing for  
CC the presence of amount of mray protein expressed in a patient sample.  
CC Mray PCR probes are useful for diagnosing diseases, and can characterise  
CC the response of the infectious organism to drugs. Mray proteins and  
CC polynucleotides are also useful for screening for antagonists, agonists  
CC and drugs against infectious micro-organisms. Mray agonists and  
CC antagonists are bacteriostatic and bacteriocidal compounds which can be  
CC used in treatment to enhance (agonist) or block (antagonist or antisense  
CC sequence) mray activity, therefore treating microbial diseases,  
CC especially Streptococcus pneumoniae diseases including otitis media,  
CC bacteraemia, conjunctivitis, pneumonia, sinusitis, pleural empyema,  
CC endocarditis and especially meningitis. Epitopes of mray proteins and  
CC polynucleotides are useful immunogens (vaccines) for producing anti-mray  
CC antibodies for prevention of bacterial infections, and mray  
CC polynucleotides can be used in genetic immunisation (gene therapy) to  
CC prevent infections. Mray proteins, polynucleotides and their  
CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and  
CC are useful for use on wounds and body implants to prevent bacterial  
CC infection.

XX  
SQ Sequence 326 AA;  
Query Match 48.0%; Score 47; DB 20; Length 326;  
Best Local Similarity 43.8%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16  
:|:|:|:|:|:|  
Db 141 YQVHLGIFYIVFALFW 156

RESULT 6  
AAM01092  
ID AAM01092 standard; Protein; 326 AA.

XX AAM01092;  
XX  
DT 02-OCT-2001 (first entry)  
XX CFE 95 protein sequence.  
DE  
KW Antibacterial; vaccine; gene therapy; bacterial cell wall viability;  
KW CFE; CEG; Conserved Essential Gene; bacterial infection;  
KW antisense therapy; antibiotic resistance.

XX Streptococcus pneumoniae.  
XX WO200149721-A2.  
PN 12-JUL-2001.  
PD  
XX 29-DEC-2000; 2000WO-US35604.  
PF  
XX 30-DEC-1999; 99US-0174089.  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PA  
XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Brucolleri RE;  
PI Thanassi JA;  
PI WPI; 2001-496721/54.  
DR N-PSDB; AAH90791.  
DR

XX Nucleic acids encoding conserved essential genes involved in bacterial  
PT replication which are potential targets for the treatment of antibiotic  
PT resistant bacterial infections -  
XX

PS Claim 27; Pages 342-343; 380pp; English.

XX  
CC The present invention relates to nucleic acids (AAH90701-AAH90918)  
CC encoding polypeptides (AAM01002-AAM01114), which are essential for the  
CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For

CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic  
CC acids are useful for detecting the presence of proteins essential for the  
CC viability of a bacterial cell wall in samples such as cells, tissues,  
CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
CC and for detecting corresponding target nucleic acid molecules with  
CC complementary sequences. The nucleic acids are also useful for  
CC determining whether a genomic nucleotide sequence of interest is  
CC essential for viability of a bacterial cell or whether it resides within  
CC an operon, by integrating an exogenous nucleotide sequence comprising a  
CC portion of an open reading frame of the genomic sequence of interest  
CC (comprising 200-500 base pairs) into the genomic sequence of interest  
CC which confers a selectable phenotype to the cell, and determining cell  
CC viability with a selection agent such as chloramphenicol. The nucleic  
CC acids and proteins are also useful as vaccines and for treating bacterial  
CC infections with gene therapy and antisense therapy. The nucleic acids  
CC also enable identification of targets suitable for the treatment of  
CC antibiotic resistant bacterial infections.

XX  
SQ Sequence 326 AA;

Query Match 48.0%; Score 47; DB 22; Length 326;  
Best Local Similarity 43.8%; Pred. No. 17;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16  
:|:|:|:|:|:|  
Db 141 YQVHLGIFYIVFALFW 156

RESULT 7  
AAM84080  
ID AAM84080 standard; Protein; 50 AA.

XX AAM84080;  
XX  
DT 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen SEQ ID NO:11673.  
DE  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.



PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-483426/52.  
DR N-PSDB; AAK56861.  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
PS Claim 11; SEQ ID NO 11673; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 50 AA;



Query Match 46.9%; Score 46; DB 22; Length 50;  
Best Local Similarity 46.2%; Pred. No. 4.1;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 NCGIDFIIFWIFW 16  
Db 12 NCCYQFLIFLYW 24

RESULT 8  
AAB57264  
ID AAB57264 standard; Protein; 64 AA.

XX AAB57264;  
XX 13-MAR-2001 (first entry)  
DE Human prostate cancer antigen protein sequence SEQ ID NO:1842.  
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease.

XX Homo sapiens.  
OS WO200055174-A1.  
PN 21-SEP-2000.  
XX 08-MAR-2000; 2000WO-US05988.

PF 12-MAR-1999; 99US-0124270.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;  
PI WPI; 2000-587513/55.  
XX N-PSDB; AAF16467.

PT Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -

PS Claim 11; Page 2304; 2338pp; English.

XX AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen  
CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.

XX Sequence 64 AA;

Query Match 46.9%; Score 46; DB 21; Length 64;  
Best Local Similarity 50.0%; Pred. No. 5.1;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 QANCGIDFIIFW 13

Db 29 ESNCGLDFVSFF 40  
::|||:|:|:

RESULT 9  
AAW54009  
ID AAW54009 standard; Protein; 123 AA.

XX AAW54009;  
XX 29-JUL-1998 (first entry)  
XX Anti-CD4 antibody VH5 monkey clone 5-11.

DE Anti-CD4 antibody; monkey; human; therapy; variable heavy domain;  
KW Old World monkey; constant domain; eczema; immuno-modulated disease;  
KW rheumatoid arthritis.

XX Primate sp.  
OS US5750105-A.  
XX 12-MAY-1998.

PD 07-JUN-1995; 95US-0476349.  
XX 10-JUL-1992; 92US-0912292.  
PR 25-JUL-1991; 91US-0735064.

PR 23-MAR-1992; 92US-0856281.  
PR 05-DEC-1995; 95US-0379072.  
PR 07-JUN-1995; 95US-0476349.

XX (IDEC-) IDEC PHARM CORP.

XX Hanna N, Newman RA, Raab RW;  
PI WPI; 1998-296690/26.

XX Improved method for antibody treatment - uses an antibody comprising  
PT an Old World monkey variable region and a human constant domain  
XX Example 1; Fig 9e; 84pp; English.

PS This sequence represents the VH5 domain of an anti-CD4 antibody

CC (Ab). This sequence can be used in the method of the invention for  
CC treating a subject, where the treatment comprises administration of an  
CC Ab. The method comprises the administration of an antibody which has an  
CC Old World monkey (e.g. baboon or macaque) variable region which binds to  
CC an antigen (Ag) (or Ag binding portion), and a human constant domain. The  
CC method is useful for the treatment of eczema and immuno-modulated  
CC diseases and especially rheumatoid arthritis. The recombinant antibodies  
CC used are sufficiently different from native monkey antibodies to allow  
CC human antigens to raise these antibodies, but similar enough to human  
CC antibody so there is no immune response to the antibodies in humans.  
CC Compared to antibodies used in therapy in prior art, these antibodies do  
CC not induce human anti-antibodies on repeated administration. They also  
CC have longer half-lives and do not have a lack of effector function with  
CC human cells.

XX Sequence 123 AA;

Query Match 46.9%; Score 46; DB 19; Length 123;  
Best Local Similarity 58.3%; Pred. No. 9.5;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16  
Db 25 CGFSFTGFWISW 36

RESULT 10  
AAM51648  
ID AAM51648 standard; Protein; 285 AA.

XX AC AAM51648;  
XX DT 20-FEB-2002 (first entry)  
XX DE Rat GPCR polypeptide.  
XX KW Rat; G-protein coupled receptor; GPCR; gene therapy;  
KW MAS proto-oncogene receptor; human protease; disease.  
XX OS Rattus sp.  
XX PN WO200181409-A2.  
XX PD 01-NOV-2001.  
XX PF 24-APR-2001; 2001WO-US13097.  
XX PR 24-APR-2000; 2000US-199149P.  
PR 04-AUG-2000; 2000US-0633146.  
XX PA (PEKE ) PE CORP NY.  
XX PI Wei M, Cravchik A, Di Francesco V, Beasley EM;  
XX WPI; 2002-049265/06.  
XX PT Novel human G protein-coupled receptor polypeptide that is related to  
PT MAS proto-oncogene receptor subfamily, useful as model and target for  
PT developing human therapeutic agent -  
XX PS Disclosure; Fig 2; 60pp; English.  
XX CC The invention relates to an isolated human G protein-coupled receptor  
CC (GPCR) polypeptide that is related to the MAS proto-oncogene receptor  
CC subfamily. The polypeptide comprises a fully defined sequence of 289  
CC amino acids as given in the specification, or its fragment comprising  
CC 10 contiguous amino acids, or an amino acid sequence of an allelic  
CC variant or orthologue of the amino acid sequence given in the sequence.  
CC The polypeptide is useful for identifying a modulator of a GPCR  
CC polypeptide or an agent that binds to it. The polypeptide is also  
CC useful for treating a disease or condition mediated by human proteases.  
CC The present sequence is a rat GPCR polypeptide used in comparison  
CC studies with the polypeptide of the invention.  
XX SQ Sequence 285 AA;  
Query Match 45.9%; Score 45; DB 23; Length 285;  
Best Local Similarity 60.0%; Pred. No. 30;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 6 GIDFIIFWIF 15  
DB 213 GIDWFLFWVF 222  
RESULT 11  
AAR48755  
ID AAR48755 standard; Protein; 298 AA.  
XX AC AAR48755;  
XX DT 07-JUN-1996 (first entry)  
XX DE Human thoracic aorta G-protein coupled receptor protein.  
XX KW G-protein coupled receptor; ligand binding assay; transmembrane domain;  
KW psychotic disorder; schizophrenia; dopamine; CAMP; adenosine; thrombin;  
KW muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;  
KW rhodopsin; opsin; odorant; cytomegalovirus.  
XX OS Homo sapiens.  
XX PI

PN WO9405695-A1.  
XX PD 17-MAR-1994.  
XX PF 09-SEP-1993; 93WO-US08528.  
XX PR 10-SEP-1992; 92US-0943236.  
XX PA (UYN Y ) UNIV NEW YORK STATE.  
XX PI Murphy RB, Schuster DI;  
XX DR WPI; 1994-101120/12.  
XX PT Polypeptides of G-coupled receptor proteins (GPRs) - useful for  
PT binding GPR ligands or modulating GPR binding  
XX PS Disclosure; Page 130-131; 160pp; English.  
XX CC Proteins AAR48685-R48758 represent a range of G-protein coupled receptor  
CC proteins selected from CAMP, adenosine, muscarinic acetylcholine,  
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,  
CC odorant, cytomegaloviral and other G-protein coupled receptors. The  
CC receptor proteins were used to design polypeptides, pref. based on the  
CC transmembrane domains, for use in G-protein coupled receptor ligand  
CC binding assays. The polypeptide fragments retain biological activity  
CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR  
CC (see AAR48759-R48758, AAR50569-R50807 and AAR9189-R89195 for examples  
CC of polypeptide fragments). The polypeptide fragments can be used in  
CC compositions for treating subjects suffering from a pathology related to  
CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia.  
XX SQ Sequence 298 AA;  
Query Match 45.9%; Score 45; DB 15; Length 298;  
Best Local Similarity 60.0%; Pred. No. 31;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 6 GIDFIIFWIF 15  
DB 211 GIDWFLFWVF 220  
RESULT 12  
AAW02727  
ID AAW02727 standard; peptide; 298 AA.  
XX AC AAW02727;  
XX DT 13-NOV-1996 (first entry)  
XX DE Human thoracic aorta G-protein coupled receptor.  
XX KW G-protein coupled receptor; ligand binding assay; transmembrane domain;  
KW schizophrenia; dopamine; CAMP; adenosine; thrombin; adrenergic; opsin;  
KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;  
KW odorant; cytomegalovirus; serotonergic.  
XX OS Homo sapiens.  
XX PN US5508384-A.  
XX PD 16-APR-1996.  
XX PF 10-SEP-1992; 92US-0943236.  
XX PR 09-SEP-1993; 93US-0118270.  
PR 10-SEP-1992; 92US-0943236.  
XX PA (UYN Y ) UNIV NEW YORK STATE.  
XX PI Murphy RB, Schuster DI;  
XX

DR WPI; 1996-208785/21.  
XX New dopamine receptor peptide - useful as antipsychotic agent, e.g.  
PT for treating schizophrenia  
PT  
XX  
PS  
XX Disclosure; Column 173-176; 184pp; English.  
PS  
CC Proteins AAW02657-W02730 represent a range of G-protein coupled receptor  
CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,  
CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,  
CC odorant, cytomegaloviral and other GPR proteins. The receptor proteins  
CC were used to design polypeptides, pref. based on the transmembrane  
CC domains, for use in G-protein coupled receptor ligand binding assays.  
CC The polypeptide fragments retain biological activity such as binding a  
CC GPR ligand or modulating GPR ligand binding to a GPR (see  
CC AAW02747-W02999 for examples of polypeptide fragments). The polypeptide  
CC fragments can be used in compositions for treating subjects suffering  
CC from a pathology related to a GPR abnormality e.g. a psychotic disorder  
CC such as schizophrenia.  
XX  
SQ Sequence 298 AA;  
Query Match 45.9%; Score 45; DB 17; Length 298;  
Best Local Similarity 60.0%; Pred. No. 31;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 6 GIDFIIIFWIF 15  
Db 211 GIDWFLFWVF 220  
|||: ||:|  
RESULT 13  
AAU76411  
ID AAU76411 standard; Protein; 323 AA.  
XX  
AC AAU76411;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Novel human G-protein coupled receptor IGPCR18.  
KW  
KW G-protein coupled; receptor; IGPCR18; cardiovascular system disorder;  
KW signal processing; female reproductive tissue; infertility;  
KW cardiovascular disease; coronary heart disease; heart attack; stroke;  
KW inflammatory disorder; metabolic disorder; uterus; placenta; ovary;  
KW prostate; reproductive disorder; pain; cancer; CNS disorder;  
KW central nervous system disorder; schizophrenia;  
KW episodic paroxysmal anxiety; EPA; obsessive compulsive disorder; COD;  
KW Parkinson's disease; multiple sclerosis; Alzheimer's disease; dementia;  
KW Huntington's disease; anorexia; kidney disease; renal failure; obesity;  
KW gastrointestinal disorder; irritable bowel syndrome; IBS; diarrhoea;  
KW motility disorder; gastric emptying; osteoporosis; infection; asthma;  
KW allergy; arthritis; sepsis.  
OS Homo sapiens.  
XX  
XX WO200202598-A2.  
PN  
PD 10-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP07530.  
PF  
XX 30-JUN-2000; 2000US-215879P.  
PR  
XX (INGE-) INGENIUM PHARM AG.  
PA  
XX Wattler F, Wattler S, Trommler P, Nehls MC;  
PI  
XX WPI; 2002-140079/18.  
DR  
DR N-PSDB; ABK15133.  
XX  
XX New human G protein-coupled receptor protein, IGPCR18, useful for  
PT diagnosis, prevention, amelioration or treatment of pain, cancer,

PT inflammatory, metabolic, reproductive, gynaecological, and  
PT cardiovascular disorders -  
XX  
PS Claim 8; Fig 2; 55pp; English.  
XX  
CC The invention describes a novel human G protein-coupled receptor (GPCR)  
CC protein, IGPCR18 (I). (I), the polynucleotide (II), agonists and  
CC antagonists of (I)/(II) are useful for the diagnosis and treatment of  
CC cardiovascular system disorders, the detection of mutant or  
CC inappropriately expressed forms of IGPCR18, and for drug screening.  
CC (I) is useful in treatment of diseases associated with signal processing  
CC in female reproductive tissues, such as infertility, cardiovascular  
CC diseases such as coronary heart disease, heart attack and stroke,  
CC inflammatory disorders and metabolic disorders linked to reproductive  
CC tissues like uterus, placenta, ovary and prostate, reproductive  
CC disorders, pain, cancer, central nervous system disorders such as  
CC schizophrenia, episodic paroxysmal anxiety (EPA) disorders such as  
CC obsessive compulsive disorder (COD), Parkinson's disease, multiple  
CC sclerosis, Alzheimer's disease/dementia, Huntington's disease, anorexia,  
CC kidney diseases such as renal failure, obesity, gastrointestinal  
CC disorders such as irritable bowel syndrome (IBS), diarrhoea, motility  
CC disorders and conditions of delayed gastric emptying, osteoporosis,  
CC infections such as bacterial, fungal, protozoal and viral infections,  
CC asthma, allergy, arthritis, and sepsis. (I) is useful for the  
CC generation of antibodies, in identification of other cellular gene  
CC products involved in regulating IGPCR18, and as pharmaceutical reagents.  
CC (II) is useful for mapping the location of the gene to the chromosome,  
CC and to detect abnormalities involving IGPCR18 gene structure. This  
CC is the amino acid sequence of the novel human G-protein coupled receptor,  
CC IGPCR18, described in the method of the invention.  
XX  
SQ Sequence 323 AA;  
Query Match 45.9%; Score 45; DB 23; Length 323;  
Best Local Similarity 60.0%; Pred. NO. 34;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 6 GIDFIIIFWIF 15  
Db 238 GIDWFLFWVF 247  
|||: ||:|  
RESULT 14  
AAE14729  
ID AAE14729 standard; Protein; 327 AA.  
XX  
AC AAE14729;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human G-protein coupled receptor, IGS43.  
KW  
KW IGS43; G-protein coupled receptor; GPCR; uterus; lung; trachea;  
KW colon; small intestine; stomach; mammary gland; prostate; testis;  
KW psychiatric disorder; central nervous system disorder; schizophrenia;  
KW episodic and paroxysmal anxiety disorder; Parkinson's disease;  
KW multiple sclerosis; Alzheimer's disease; kidney disease; cardiovascular disease;  
KW heart failure; angina pectoris; kidney disease; dyslipidaemias; obesity;  
KW emesis; gastrointestinal disorder; inflammatory bowel disease; diabetes;  
KW osteoporosis; inflammation; infection; human immunodeficiency virus; HIV;  
KW cancer; immune disorder; urinary retention; asthma; allergy; arthritis;  
KW benign prostatic hypertrophy; endotoxin shock; sepsis; gene therapy;  
KW gynaecological disorder; vaccine; human.  
XX  
XX Homo sapiens.  
OS  
XX WO200228897-A2.  
PN  
XX 11-APR-2002.  
PD  
XX 28-SEP-2001; 2001WO-EP11319.  
PF  
XX 02-OCT-2000; 2000EP-0203411.  
PR

PR 04-OCT-2000; 2000US-237394P.  
XX (SOLV ) SOLVAY PHARM BV.  
PA  
XX Deleersnijder W, Blockx H, De Moor L;  
PI  
XX WPI; 2002-426102/45.  
DR N-PSDB; AAD31159.  
DR  
XX Novel G-protein coupled receptor, termed IGS43 polypeptide and nucleic  
PT acid encoding the polypeptide, useful for treating disorders of uterus,  
PT kidney, lung, colon, stomach, mammary gland, prostate and testis -  
XX  
PS Claim 25; Page 7; 59pp; English.  
XX  
CC The invention relates to IGS43 G-protein coupled receptor (GPCR) and the  
CC polynucleotide encoding it. The IGS43 polypeptide, polynucleotide,  
CC agonist, antagonist or antibody is useful for treating dysfunctions or  
CC disorders related to uterus, kidney, lung, trachea, colon, small  
CC intestine, stomach, mammary gland, prostate, testis, central nervous  
CC system, cerebellum and spinal cord. The polypeptide is useful as vaccine  
CC for inducing immunological response in a mammal, for treating  
CC psychiatriac and central nervous system disorders including  
CC schizophrenia, episodic and paroxysmal anxiety disorders e.g. obsessive  
CC compulsive disorder, post traumatic disorder, phobia and panic, major  
CC depressive disorder, bipolar disorder, Parkinson's disease, general  
CC anxiety disorder, autism, delirium, multiple sclerosis, Alzheimer's  
CC disease, dementia, severe mental retardation, Huntington's disease,  
CC dyskinesias, Tourette's syndrome, tics, tremor, dystonia, spasms,  
CC anorexia, bulimia, stroke, addiction/dependency/craving, sleep  
CC disorder, epilepsy, migraine, attention deficit/hyperactivity disorder,  
CC cardiovascular diseases including heart failure, angina pectoris,  
CC arrhythmias, myocardial infarction, cardiac hypertrophy, hypotension,  
CC hypertension, thrombosis, arteriosclerosis, cerebral vasospasm,  
CC subarachnoid haemorrhage, cerebral ischaemia, cerebral infarction,  
CC peripheral vascular disease, Raynaud's disease, kidney disease,  
CC dyslipidaemias, obesity, emesis, gastrointestinal disorders including  
CC irritable bowel syndrome, inflammatory bowel disease, diabetic  
CC gastroparesis and diabetes, ulcers, diarrhoea, osteoporosis,  
CC inflammations, infections including bacterial, fungal, protozoan and  
CC viral infections, particularly human immunodeficiency virus (HIV)-1 or  
CC HIV-2 infections, pain, cancers, chemotherapy induced injury, tumour  
CC invasion, immune disorders, urinary retention, asthma, allergies,  
CC arthritis, benign prostatic hypertrophy, endotoxin shock, sepsis,  
CC complications of diabetes mellitus, and gynaecological disorders.  
CC The polypeptide and polynucleotide of the invention are also useful as  
CC research reagents and materials for discovery of treatments and  
CC diagnostics to animal and human diseases. The polynucleotide is  
CC also useful for chromosome identification. The polypeptide is also useful  
CC for assessing the binding of small molecule substrates and ligands in  
CC cells, cell-free preparations, chemical libraries and natural product  
CC mixtures. The present sequence is human IGS43 GPCR.  
XX  
SQ Sequence 327 AA;

Query Match 45.9%; Score 45; DB 23; Length 327;  
Best Local Similarity 60.0%; Pred. No. 34;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15  
Db 240 GIDWFLFWVF 249  
|||: ||:|

RESULT 15  
AAR96145  
ID AAR96145 standard; Protein; 343 AA.  
XX  
AC AAR96145;  
XX  
DT 29-OCT-1996 (first entry)  
XX  
DE G protein coupled receptor protein.

XX G protein coupled receptor protein; GCR; identification; detection;  
KW ligands; physiological response; arachidonic acid; acetylcholine;  
KW calcium; cyclic adenosine monophosphate; cAMP; inositol;  
KW membrane potential; phosphorylation; c-fos; antibodies;  
KW transgenic animals.  
XX  
OS Oryctolagus cuniculus.  
XX  
PN EP711831-A2.  
XX  
PD 15-MAY-1996.  
XX  
PF 11-NOV-1995; 95EP-0117786.  
XX  
PR 24-AUG-1995; 95JP-0215798.  
PR 14-NOV-1994; 94JP-0279545.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
PI Fuji R, Hinuma S, Kawamata Y;  
XX  
DR WPI; 1996-232096/24.  
DR N-PSDB; AAT27218.  
XX  
PT New isolated rabbit G protein receptor protein - used partic. to  
PT identify agonists or antagonists which can promote or inhibit  
PT physiological responses  
XX  
PS Claim 1; Page 30-32; 40pp; English.  
XX  
CC G protein coupled receptor (GCR) proteins can be used to identify  
CC ligands which act as antagonists of GCR/ligand binding. Such ligands  
CC can be used for promoting or inhibiting physiological responses such  
CC as liberation of arachidonic acid, acetylcholine and endocellular  
CC calcium, endocellular cyclic AMP production, production of inositol,  
CC changes in cell membrane potential, phosphorylation of endocellular  
CC proteins, activation of c-fos, lowering of pH, activation of G  
CC protein and cell promulgation. The GCR products can also be used for  
CC the detection of ligands and in the production of antibodies and  
CC transgenic animals.  
XX  
SQ Sequence 343 AA;

Query Match 45.9%; Score 45; DB 17; Length 343;  
Best Local Similarity 60.0%; Pred. No. 36;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15  
Db 256 GIDWFLFWVF 265  
|||: ||:|

Search completed: July 9, 2003, 15:07:41  
Job time : 9.92035 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 4.95575 Seconds  
(without alignments)  
94.994 Million cell updates/sec

Title: US-09-854-133-587

Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/PCTvs\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score		Query Match		Length	DB ID	Description
1	46	46.9	123	1	US-08-478-039-82	Sequence 82, Appl	
2	46	46.9	123	1	US-08-476-349A-82	Sequence 82, Appl	
3	45	45.9	298	1	US-08-118-270-76	Sequence 76, Appl	
4	45	45.9	298	5	PCT-US93-08528-76	Sequence 76, Appl	
5	42	42.9	125	2	US-08-665-202-45	Sequence 45, Appl	
6	42	42.9	125	2	US-08-665-202-46	Sequence 46, Appl	
7	42	42.9	125	2	US-08-665-202-49	Sequence 49, Appl	
8	42	42.9	125	2	US-08-665-202-51	Sequence 51, Appl	
9	42	42.9	342	2	US-08-483-151-2	Sequence 2, Appl	
10	42	42.9	518	4	US-09-134-001C-4069	Sequence 4069, Ap	
11	41	41.8	119	2	US-08-318-157B-2	Sequence 2, Appl	
12	41	41.8	119	2	US-08-318-157B-12	Sequence 12, Appl	
13	41	41.8	119	2	US-08-318-157B-17	Sequence 17, Appl	
14	41	41.8	125	2	US-08-665-202-48	Sequence 48, Appl	
15	41	41.8	125	2	US-08-665-202-50	Sequence 50, Appl	
16	41	41.8	125	2	US-08-665-202-52	Sequence 52, Appl	
17	41	41.8	125	2	US-08-665-202-53	Sequence 53, Appl	
18	41	41.8	125	2	US-08-665-202-54	Sequence 54, Appl	
19	41	41.8	125	2	US-08-665-202-55	Sequence 55, Appl	
20	41	41.8	125	2	US-08-665-202-57	Sequence 57, Appl	
21	41	41.8	139	4	US-09-136-315-2	Sequence 2, Appl	
22	41	41.8	139	4	US-09-136-315-6	Sequence 6, Appl	
23	40	40.8	119	2	US-08-318-157B-8	Sequence 8, Appl	
24	40	40.8	119	2	US-08-318-157B-9	Sequence 9, Appl	
25	40	40.8	119	2	US-08-318-157B-10	Sequence 10, Appl	
26	40	40.8	119	2	US-08-318-157B-11	Sequence 11, Appl	
27	40	40.8	119	2	US-08-318-157B-13	Sequence 13, Appl	

28	40	40.8	119	2	US-08-318-157B-14	Sequence 14, Appl
29	40	40.8	119	2	US-08-318-157B-15	Sequence 15, Appl
30	40	40.8	203	4	US-09-124-141-13	Sequence 13, Appl
31	40	40.8	203	4	US-09-124-141-21	Sequence 21, Appl
32	40	40.8	593	4	US-09-124-141-7	Sequence 7, Appl
33	40	40.8	594	2	US-08-592-696-4	Sequence 4, Appl
34	40	40.8	594	2	US-09-027-536-4	Sequence 4, Appl
35	40	40.8	594	3	US-09-028-148-4	Sequence 4, Appl
36	40	40.8	594	4	US-09-124-141-15	Sequence 15, Appl
37	40	40.8	594	4	US-09-124-141-23	Sequence 23, Appl
38	40	40.8	625	1	US-08-242-689-2	Sequence 2, Appl
39	40	40.8	625	4	US-08-468-583-2	Sequence 2, Appl
40	39.5	40.3	1269	3	US-08-781-891-74	Sequence 74, Appl
41	39	39.8	36	2	US-08-576-626A-44	Sequence 44, Appl
42	39	39.8	114	3	US-08-483-749A-10	Sequence 10, Appl
43	39	39.8	262	1	US-08-480-882B-5	Sequence 5, Appl
44	39	39.8	262	1	US-08-480-210-5	Sequence 5, Appl
45	39	39.8	262	1	US-08-220-401-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-478-039-82  
; Sequence 82, Application US/08478039  
; Patent No. 5681722  
; GENERAL INFORMATION:  
; APPLICANT: Newman, Roland A.  
; APPLICANT: Hanna, Nabil  
; APPLICANT: Raab, Ronald W.  
; TITLE OF INVENTION: Recombinant Antibodies for Human Therapy  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince St.  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,039  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/379,072  
; FILING DATE: 25-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/912,292  
; FILING DATE: 10-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/856,281  
; FILING DATE: 23-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/735,064  
; FILING DATE: 25-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Teskin Esq., Robin L.  
; REGISTRATION NUMBER: 35,030  
; REFERENCE/DOCKET NUMBER: 012712-160  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6620  
; TELEFAX: 703-836-2021  
; INFORMATION FOR SEQ ID NO: 82:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 123 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant

;; TOPOLOGY: not relevant  
;; MOLECULE TYPE: peptide  
;; ORIGINAL SOURCE:  
;; ORGANISM: Monkey  
;; POSITION IN GENOME:  
;; CHROMOSOME/SEGMENT: VH5 clone 5-11  
US-08-478-039-82

Query Match 46.9%; Score 46; DB 1; Length 123;  
Best Local Similarity 58.3%; Pred. No. 3.9;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 CGIDFIIFWIFW 16  
|| | |||  
Db 25 CGFSFTGFWISW 36

## RESULT 2

US-08-476-349A-82  
; Sequence 82, Application US/08476349A  
; Patent No. 5750105

## ; GENERAL INFORMATION:

;; APPLICANT: Newman, Roland A.  
;; APPLICANT: Hanna, Nabil  
;; APPLICANT: Raab, Ronald W.  
;; TITLE OF INVENTION: Recombinant Antibodies for Human Therapy  
;; NUMBER OF SEQUENCES: 114  
;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
;; STREET: 699 Prince St.  
;; CITY: Alexandria  
;; STATE: VA

;; COUNTRY: USA

;; ZIP: 22313-1404

## ; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/476,349A  
;; FILING DATE: 07-JUN-1995

;; CLASSIFICATION: 514

## ; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/379,072

;; FILING DATE: 25-JAN-1995

## ; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/912,292

;; FILING DATE: 10-JUL-1992

## ; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/856,281

;; FILING DATE: 23-MAR-1992

## ; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/735,064

;; FILING DATE: 25-JUL-1991

## ; ATTORNEY/AGENT INFORMATION:

;; NAME: teskin Esq., Robin L.

;; REGISTRATION NUMBER: 35,030

;; REFERENCE/DOCKET NUMBER: 012712-161

## ; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 703-836-6620

;; TELEFAX: 703-836-2021

## ; INFORMATION FOR SEQ ID NO: 82:

## ; SEQUENCE CHARACTERISTICS:

;; LENGTH: 123 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: not relevant

;; TOPOLOGY: not relevant

;; MOLECULE TYPE: peptide

## ; ORIGINAL SOURCE:

;; ORGANISM: Monkey

;; POSITION IN GENOME:

;; CHROMOSOME/SEGMENT: VH5 clone 5-11

US-08-476-349A-82

Query Match 46.9%; Score 46; DB 1; Length 123;  
Best Local Similarity 58.3%; Pred. No. 3.9;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 CGIDFIIFWIFW 16  
|| | |||  
Db 25 CGFSFTGFWISW 36

## RESULT 3

US-08-118-270-76  
; Sequence 76, Application US/08118270  
; Patent No. 5508384

## ; GENERAL INFORMATION:

;; APPLICANT: Murphy, Randall B.

;; APPLICANT: Schuster, David I.

;; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN

;; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF

;; NUMBER OF SEQUENCES: 348

## ; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: BROWDY AND NEIMARK

;; STREET: 419 Seventh Street, N.W., Suite 300

;; CITY: Washington

;; STATE: D.C.

;; COUNTRY: USA

;; ZIP: 20004

## ; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25

## ; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/118,270

;; FILING DATE: 09-SEP-1993

## ; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/943,236

;; FILING DATE: 10-SEP-1992

## ; ATTORNEY/AGENT INFORMATION:

;; NAME: Townsend, Kevin G.

;; REGISTRATION NUMBER: 34,033

;; REFERENCE/DOCKET NUMBER: MURPHY-2A

## ; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 202-628-5197

;; TELEFAX: 202-737-3528

;; TELEX: 248633

## ; INFORMATION FOR SEQ ID NO: 76:

## ; SEQUENCE CHARACTERISTICS:

;; LENGTH: 298 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: peptide

US-08-118-270-76

Query Match 45.9%; Score 45; DB 1; Length 298;  
Best Local Similarity 60.0%; Pred. No. 14;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 6 GIDFIIFWIF 15  
|||: |||:  
Db 211 GIDWFLFWVF 220

## RESULT 4

PCT-US93-08528-76

; Sequence 76, Application PC/TUS9308528

## ; GENERAL INFORMATION:

;; APPLICANT: New York University

;; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN

;; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF

;; NUMBER OF SEQUENCES: 348

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BROWDY AND NEIMARK  
;; STREET: 419 Seventh Street, N.W., Suite 300  
;; CITY: Washington  
;; STATE: D.C.  
;; COUNTRY: USA  
;; ZIP: 20004  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/08528  
;; FILING DATE: 09-SEP-1993  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/943,236  
;; FILING DATE: 10-SEP-1992  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Townsend, Kevin G.  
;; REGISTRATION NUMBER: 34,033  
;; REFERENCE/DOCKET NUMBER: MURPHY=2 PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-628-5197  
;; TELEFAX: 202-737-3528  
;; TELEX: 248633  
;;  
;; INFORMATION FOR SEQ ID NO: 76:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 298 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; PCT-US93-08528-76  
;;  
Query Match 45.9%; Score 45; DB 5; Length 298;  
Best Local Similarity 60.0%; Pred. No. 14;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 6 GIDFIIFWIF 15  
| | | : | | |  
Db 211 GIDWFLFWVF 220  
  
RESULT 5  
US-08-665-202-45  
; Sequence 45, Application US/08665202  
; Patent No. 5977322  
; GENERAL INFORMATION:  
; APPLICANT: Marks, James D.  
; APPLICANT: Schier, Robert  
; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to  
; TUMOR ANTIGENS  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,202  
; FILING DATE: 13-JUN-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,202  
; FILING DATE: 13-JUN-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/000,238  
; FILING DATE: 14-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 02307E-061410  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 125 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/000,250  
;; FILING DATE: 15-JUN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hunter, Tom  
;; REGISTRATION NUMBER: 38,498  
;; REFERENCE/DOCKET NUMBER: 02307E-061410  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 576-0200  
;; TELEFAX: (415) 576-0300  
;; INFORMATION FOR SEQ ID NO: 45:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 125 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; US-08-665-202-45  
;;  
Query Match 42.9%; Score 42; DB 2; Length 125;  
Best Local Similarity 54.5%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 6 GIDFIIFWIF 16  
| | | : | | |  
Db 26 GYDFTTYWIAW 36  
  
RESULT 6  
US-08-665-202-46  
; Sequence 46, Application US/08665202  
; Patent No. 5977322  
; GENERAL INFORMATION:  
; APPLICANT: Marks, James D.  
; APPLICANT: Schier, Robert  
; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to  
; TUMOR ANTIGENS  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,202  
; FILING DATE: 13-JUN-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/000,238  
; FILING DATE: 14-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 02307E-061410  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 125 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear

; MOLECULE TYPE: peptide  
US-08-665-202-46

Query Match 42.9%; Score 42; DB 2; Length 125;  
Best Local Similarity 54.5%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 GIDFIIFWIFW 16  
| | | : | | |  
Db 26 GYDFTTYWIAW 36

## RESULT 7

US-08-665-202-49  
; Sequence 49, Application US/08665202  
; Patent No. 5977322

; GENERAL INFORMATION:

; APPLICANT: Marks, James D.  
; APPLICANT: Schier, Robert  
; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to  
; Tumor Antigens  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/665,202  
; FILING DATE: 13-JUN-1996  
; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,238  
; FILING DATE: 14-JUN-1995  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,250  
; FILING DATE: 15-JUN-1995  
; ATTORNEY/AGENT INFORMATION:

; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 02307E-061410

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 125 amino acids  
; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide  
US-08-665-202-49

Query Match 42.9%; Score 42; DB 2; Length 125;  
Best Local Similarity 54.5%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 GIDFIIFWIFW 16  
| | | : | | |  
Db 26 GYDFTTYWIAW 36

## RESULT 8

US-08-665-202-51  
; Sequence 51, Application US/08665202  
; Patent No. 5977322

; GENERAL INFORMATION:

; APPLICANT: Marks, James D.

; APPLICANT: Schier, Robert

; TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to  
; Tumor Antigens

; NUMBER OF SEQUENCES: 141

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/665,202

; FILING DATE: 13-JUN-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,238

; FILING DATE: 14-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/000,250

; FILING DATE: 15-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Hunter, Tom

; REGISTRATION NUMBER: 38,498

; REFERENCE/DOCKET NUMBER: 02307E-061410

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 51:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 125 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-665-202-51

Query Match 42.9%; Score 42; DB 2; Length 125;  
Best Local Similarity 54.5%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 GIDFIIFWIFW 16  
| | | : | | |  
Db 26 GYDFTTYWIAW 36

## RESULT 9

US-08-483-151-2  
; Sequence 2, Application US/08483151  
; Patent No. 5858752

; GENERAL INFORMATION:

; APPLICANT: Seed, Brian

; APPLICANT: Holgersson, Jan

; TITLE OF INVENTION: FUCOSYLTRANSFERASE GENES AND USES THEREOF

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS



; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,151  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lech, Karen F.  
; REGISTRATION NUMBER: 35,238  
; REFERENCE/DOCKET NUMBER: 00786/278001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 342 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-483-151-2

Query Match 42.9%; Score 42; DB 2; Length 342;  
Best Local Similarity 45.5%; Pred. No. 45;  
Matches 5; Conservative 2; Mismatches 4; Indels 4; Gaps 0;

QY 6 GIDFIIFWIFW 16  
| | | | |  
Db 21 GATFMVWIFW 31

## RESULT 10

US-09-134-001C-4069  
; Sequence 4069, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4069  
; LENGTH: 518  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (6)  
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
US-09-134-001C-4069

Query Match 42.9%; Score 42; DB 4; Length 518;  
Best Local Similarity 58.3%; Pred. No. 69;  
Matches 7; Conservative 2; Mismatches 3; Indels 3; Gaps 0;

QY 2 QANCGIDFIIFW 13  
: | | | | | | |  
Db 175 RSNCGIGFKGF 186

## RESULT 11

US-08-318-157B-2  
; Sequence 2, Application US/08318157B  
; Patent No. 5874540  
; GENERAL INFORMATION:

; APPLICANT: HANSEN, Hans J.  
; APPLICANT: ARMOUR, Kathryn L.  
; TITLE OF INVENTION: CDR-GRAFTED TYPE III ANTI-CEA HUMANIZED

; TITLE OF INVENTION: MOUSE MONOCLONAL ANTIBODIES  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/318,157B  
; FILING DATE: 05-OCT-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SAXE, Bernhard D.  
; REGISTRATION NUMBER: 28,665  
; REFERENCE/DOCKET NUMBER: 18733/464  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 119 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-318-157B-2

Query Match 41.8%; Score 41; DB 2; Length 119;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 ANCGIDFIIFWIFW 16  
| | | | | | |  
Db 23 AASGDFTTYWMSW 36

## RESULT 12

US-08-318-157B-12  
; Sequence 12, Application US/08318157B  
; Patent No. 5874540  
; GENERAL INFORMATION:  
; APPLICANT: HANSEN, Hans J.  
; APPLICANT: ARMOUR, Kathryn L.  
; TITLE OF INVENTION: CDR-GRAFTED TYPE III ANTI-CEA HUMANIZED  
; TITLE OF INVENTION: MOUSE MONOCLONAL ANTIBODIES  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/318,157B  
; FILING DATE: 05-OCT-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SAXE, Bernhard D.  
; REGISTRATION NUMBER: 28,665  
; REFERENCE/DOCKET NUMBER: 18733/464

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-318-157B-12

Query Match 41.8%; Score 41; DB 2; Length 119;  
Best Local Similarity 35.7%; Pred. No. 21;  
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 ANCGIDFIIFWIFW 16  
::|||:|:  
Db 23 SSSGFDFTTYWMSW 36

RESULT 13  
US-08-318-157B-17  
Sequence 17, Application US/08318157B  
Patent No. 5874540  
GENERAL INFORMATION:  
APPLICANT: HANSEN, Hans J.  
APPLICANT: ARMOUR, Kathryn L.  
TITLE OF INVENTION: CDR-GRAFTED TYPE III ANTI-CEA HUMANIZED  
TITLE OF INVENTION: MOUSE MONOCLONAL ANTIBODIES  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,157B  
FILING DATE: 05-OCT-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 18733/464  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-318-157B-17

Query Match 41.8%; Score 41; DB 2; Length 119;  
Best Local Similarity 35.7%; Pred. No. 21;  
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 ANCGIDFIIFWIFW 16  
::|||:|:  
Db 23 SSSGFDFTTYWMSW 36

RESULT 14  
US-08-665-202-48  
Sequence 48, Application US/08665202  
Patent No. 5977322  
GENERAL INFORMATION:  
APPLICANT: Marks, James D.  
APPLICANT: Schier, Robert  
TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to  
TITLE OF INVENTION: Tumor Antigens  
NUMBER OF SEQUENCES: 141  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,202  
FILING DATE: 13-JUN-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/000,238  
FILING DATE: 14-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/000,250  
FILING DATE: 15-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 02307E-061410  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 125 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-665-202-48

Query Match 41.8%; Score 41; DB 2; Length 125;  
Best Local Similarity 54.5%; Pred. No. 22;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 GIDFIIFWIFW 16  
|||:|:  
Db 26 GYDFSTYWIW 36

RESULT 15  
US-08-665-202-50  
Sequence 50, Application US/08665202  
Patent No. 5977322  
GENERAL INFORMATION:  
APPLICANT: Marks, James D.  
APPLICANT: Schier, Robert  
TITLE OF INVENTION: No. 5977322el High Affinity Human Antibodies to  
TITLE OF INVENTION: Tumor Antigens  
NUMBER OF SEQUENCES: 141  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA

; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,202  
; FILING DATE: 13-JUN-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/000,238  
; FILING DATE: 14-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/000,250  
; FILING DATE: 15-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 02307E-061410  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 125 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-665-202-50

Query Match 41.8%; Score 41; DB 2; Length 125;  
Best Local Similarity 54.5%; Pred. NO. 22;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 6 GIDFIIFWIFW 16  
| | | : | |  
Db 26 GYDFSTYWIAW 36

Search completed: July 9, 2003, 15:08:27  
Job time : 5.95575 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:06:38 ; Search time 5.38053 Seconds  
(without alignments)  
346.196 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published\_Applications\_AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	100.0	16	9	US-09-854-133-587
2	98	100.0	16	9	US-10-144-649A-587
3	98	100.0	16	10	US-09-738-973-587
4	98	100.0	97	9	US-09-854-133-586
5	98	100.0	97	9	US-10-144-649A-586
6	98	100.0	97	10	US-09-738-973-586
7	98	100.0	114	9	US-10-144-649A-742
8	46	46.9	64	10	US-09-925-300-1842
9	46	46.9	123	10	US-09-850-165-90
10	45	45.9	323	9	US-10-266-643-4
11	45	45.9	323	10	US-09-816-087-4
12	45	45.9	343	9	US-10-176-079-2
13	45	45.9	343	9	US-10-176-079-9
14	45	45.9	343	9	US-09-929-752-2
15	45	45.9	343	9	US-09-929-752-9
16	45	45.9	343	9	US-10-225-567A-482
17	45	45.9	343	9	US-10-184-426-2
18	45	45.9	343	9	US-10-184-426-9
19	45	45.9	343	10	US-09-985-694A-2

20	45	45.9	343	10	US-09-985-694A-9	Sequence 9, Appli
21	43	43.9	248	9	US-09-925-299-973	Sequence 973, App
22	43	43.9	248	10	US-09-925-299-973	Sequence 973, App
23	43	43.9	475	10	US-09-815-242-11503	Sequence 11503, A
24	43	43.9	519	9	US-09-895-913A-118	Sequence 118, App
25	43	43.9	519	10	US-09-815-242-11388	Sequence 11388, A
26	42.5	43.4	393	10	US-09-784-077-2	Sequence 2, Appli
27	42	42.9	42	10	US-09-789-561-132	Sequence 132, App
28	42	42.9	619	10	US-09-729-094-4	Sequence 4, Appli
29	41	41.8	77	9	US-10-091-504-1073	Sequence 1073, Ap
30	41	41.8	77	10	US-09-764-869-1073	Sequence 1073, Ap
31	41	41.8	119	10	US-09-253-794-2	Sequence 2, Appli
32	41	41.8	119	10	US-09-253-794-12	Sequence 12, Appli
33	41	41.8	119	10	US-09-253-794-17	Sequence 17, Appli
34	41	41.8	230	9	US-09-738-626-5746	Sequence 5746, Ap
35	41	41.8	310	9	US-09-804-291-485	Sequence 485, App
36	41	41.8	310	9	US-10-183-116-51	Sequence 51, Appli
37	41	41.8	310	10	US-09-886-055-485	Sequence 485, App
38	41	41.8	676	9	US-10-295-403-100	Sequence 100, App
39	40	40.8	71	10	US-09-925-302-881	Sequence 881, App
40	40	40.8	119	10	US-09-253-794-8	Sequence 8, Appli
41	40	40.8	119	10	US-09-253-794-9	Sequence 9, Appli
42	40	40.8	119	10	US-09-253-794-10	Sequence 10, Appli
43	40	40.8	119	10	US-09-253-794-11	Sequence 11, Appli
44	40	40.8	119	10	US-09-253-794-13	Sequence 13, Appli
45	40	40.8	119	10	US-09-253-794-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1  
US-09-854-133-587  
; Sequence 587, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 587  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-587

Query Match 100.0%; Score 98; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FQANCGIDFIIFWIFW 16  
Db 1 FQANCGIDFIIFWIFW 16

RESULT 2  
US-10-144-649A-587  
; Sequence 587, Application US/10144649A  
; Publication No. US20030118599A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Fan, Ligu  
; APPLICANT: Algate, Paul A.  
; APPLICANT: McNeill, Patricia D.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C11  
; CURRENT APPLICATION NUMBER: US/10/144,649A  
; CURRENT FILING DATE: 2002-08-21  
; NUMBER OF SEQ ID NOS: 749  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 587  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-144-649A-587

Query Match 100.0%; Score 98; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 1 FQANCGIDFIIFWIFW 16

RESULT 3  
US-09-738-973-587  
; Sequence 587, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliott, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSEQ for Windows Version 3.0.  
; SEQ ID NO 587  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-973-587

Query Match 100.0%; Score 98; DB 10; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 1 FQANCGIDFIIFWIFW 16

RESULT 4  
US-09-854-133-586  
; Sequence 586, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 586  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-854-133-586

Query Match 100.0%; Score 98; DB 9; Length 97;  
Best Local Similarity 100.0%; Pred. No. 7.8e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 35 FQANCGIDFIIFWIFW 50

RESULT 5  
US-10-144-649A-586  
; Sequence 586, Application US/10144649A  
; Publication No. US20030118599A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Fan, Liqun  
; APPLICANT: Algate, Paul A.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C11  
; CURRENT APPLICATION NUMBER: US/10/144,649A  
; CURRENT FILING DATE: 2002-08-21  
; NUMBER OF SEQ ID NOS: 749  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 586  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-144-649A-586

Query Match 100.0%; Score 98; DB 9; Length 97;  
Best Local Similarity 100.0%; Pred. No. 7.8e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FQANCGIDFIIFWIFW 16  
| | | | | | | | | | | | | | | |  
Db 35 FQANCGIDFIIFWIFW 50

RESULT 6  
US-09-738-973-586  
; Sequence 586, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indirias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliott, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973

```

RESULT 10
US-10-266-643-4
; Sequence 4, Application US/102666643
; Publication No. US20030059891A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui
; TITLE OF INVENTION: ISOLATED HUMAN
; TITLE OF INVENTION: RECEPTORS, AND
; TITLE OF INVENTION: PROTEINS, AND
; FILE REFERENCE: CL000749-CON
; CURRENT APPLICATION NUMBER: US/10/
; CURRENT FILING DATE: 2002-10-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Vers
; SEQ ID NO 4
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-266-643-4

```

Query Match 45.9%; Score 45; DB 9; Length 323;  
Best Local Similarity 60.0%; Pred. No. 25;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIIFWIF 15  
|||: |||:|  
Db 244 GIDWFLFWVF 253

RESULT 11  
US-09-816-087-4  
; Sequence 4, Application US/09816087  
; Patent No. US20020064822A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui  
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED  
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR  
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000749-CIP  
; CURRENT APPLICATION NUMBER: US/09/816,087  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-816-087-4

Query Match 45.9%; Score 45; DB 10; Length 323;  
Best Local Similarity 60.0%; Pred. No. 25;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIIFWIF 15  
|||: |||:|  
Db 244 GIDWFLFWVF 253

RESULT 12  
US-10-176-079-2  
; Sequence 2, Application US/10176079  
; Publication No. US20020192760A1  
; GENERAL INFORMATION:  
; APPLICANT: Li et al.  
; TITLE OF INVENTION: G-Protein Coupled Receptor  
; FILE REFERENCE: PF145P1D1C1  
; CURRENT APPLICATION NUMBER: US/10/176,079  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: 09/562,909  
; PRIOR FILING DATE: 2000-05-02  
; PRIOR APPLICATION NUMBER: 08/461,989  
; PRIOR FILING DATE: 1995-06-05  
; PRIOR APPLICATION NUMBER: PCT/US94/13296  
; PRIOR FILING DATE: 1994-11-18  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 343  
; TYPE: PRT  
; ORGANISM: human  
US-10-176-079-2

Query Match 45.9%; Score 45; DB 9; Length 343;  
Best Local Similarity 60.0%; Pred. No. 27;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIIFWIF 15  
|||: |||:|  
Db 256 GIDWFLFWVF 265

RESULT 13  
US-10-176-079-9

; Sequence 9, Application US/10176079  
; Publication No. US20020192760A1  
; GENERAL INFORMATION:  
; APPLICANT: Li et al.  
; TITLE OF INVENTION: G-Protein Coupled Receptor  
; FILE REFERENCE: PF145P1D1C1  
; CURRENT APPLICATION NUMBER: US/10/176,079  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: 09/562,909  
; PRIOR FILING DATE: 2000-05-02  
; PRIOR APPLICATION NUMBER: 08/461,989  
; PRIOR FILING DATE: 1995-06-05  
; PRIOR APPLICATION NUMBER: PCT/US94/13296  
; PRIOR FILING DATE: 1994-11-18  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 9  
; LENGTH: 343  
; TYPE: PRT  
; ORGANISM: human  
US-10-176-079-9

Query Match 45.9%; Score 45; DB 9; Length 343;  
Best Local Similarity 60.0%; Pred. No. 27;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIIFWIF 15  
|||: |||:|  
Db 256 GIDWFLFWVF 265

RESULT 14  
US-09-929-752-2  
; Sequence 2, Application US/09929752  
; Publication No. US20030113909A1  
; GENERAL INFORMATION:  
; APPLICANT: Hinuma, Shuji  
; Fujii, Ryo  
; Kawamata, Yuji  
; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,  
; PRODUCTION AND USE THEREOF  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/929,752  
; FILING DATE: 14-Aug-2001  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/555,905  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: JP 7-215798  
; FILING DATE: 24-AUG-1995  
; APPLICATION NUMBER: JP 6-326611  
; FILING DATE: 28-DEC-1994  
; APPLICATION NUMBER: JP 7-007177  
; FILING DATE: 20-JAN-1995  
; APPLICATION NUMBER: JP 7-057186  
; FILING DATE: 16-MAR-1995  
; APPLICATION NUMBER: JP 7-224544  
; FILING DATE: 10-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Conlin, David G.  
; REGISTRATION NUMBER: 27,026



REFERENCE/DOCKET NUMBER: 1550/45836  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 343 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-929-752-2

Query Match 45.9%; Score 45; DB 9; Length 343;  
Best Local Similarity 60.0%; Pred. No. 27;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15  
|||: |||:|  
Db 256 GIDWFLFWVF 265

RESULT 15

US-09-929-752-9  
Sequence 9, Application US/09929752  
Publication No. US20030113909A1  
GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji

Fujii, Ryo

Kawamata, Yuji

TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,  
PRODUCTION AND USE THEREOF

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/929,752

FILING DATE: 14-Aug-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/555,905

FILING DATE: <Unknown>

APPLICATION NUMBER: JP 7-215798

FILING DATE: 24-AUG-1995

APPLICATION NUMBER: JP 6-326611

FILING DATE: 28-DEC-1994

APPLICATION NUMBER: JP 7-007177

FILING DATE: 20-JAN-1995

APPLICATION NUMBER: JP 7-057186

FILING DATE: 16-MAR-1995

APPLICATION NUMBER: JP 7-224544

FILING DATE: 10-AUG-1995

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 1550/45836

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 343 amino acids

TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-09-929-752-9

Query Match 45.9%; Score 45; DB 9; Length 343;  
Best Local Similarity 60.0%; Pred. No. 27;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15  
|||: |||:|  
Db 256 GIDWFLFWVF 265

Search completed: July 9, 2003, 15:10:29  
Job time : 6.38053 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 7.22124 Seconds  
(without alignments)  
213.004 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	51.0	544	2 S61985	ALG6 protein - yea
2	48	49.0	262	1 JQ1724	E1 membrane glycop
3	47	48.0	491	2 D95366	NADH2 dehydrogenas
4	45	45.9	343	2 A35639	G protein-coupled
5	44	44.9	179	1 D69890	conserved hypothet
6	44	44.9	391	2 B96994	Fe-S oxidoreductas
7	43	43.9	284	2 AH0415	probable membrane-
8	43	43.9	332	2 B81252	NADH2 dehydrogenas
9	43	43.9	475	2 A71937	amino acid permeas
10	43	43.9	519	1 A64647	amino acid permeas
11	43	43.9	545	2 B89823	hypothetical prote
12	43	43.9	804	1 YUMU	sucrose synthase (
13	42	42.9	121	2 A30560	Ig heavy chain V r
14	42	42.9	147	2 S64252	probable membrane
15	42	42.9	194	2 AB2320	hypothetical prote
16	42	42.9	256	2 T05554	hypothetical prote
17	42	42.9	329	1 C69516	hypothetical prote
18	42	42.9	483	2 S41689	cytochrome-c oxida
19	41	41.8	123	2 PH1423	Ig heavy chain V r
20	41	41.8	127	2 PH1409	Ig heavy chain V r
21	41	41.8	130	2 S06817	Ig heavy chain V r
22	41	41.8	134	2 PH1422	Ig heavy chain V r
23	41	41.8	194	2 T11071	ATPase subunit 6 -
24	41	41.8	216	2 E75425	hypothetical prote
25	41	41.8	342	2 B64395	malic acid transpo
26	41	41.8	345	2 T25580	hypothetical prote
27	41	41.8	495	1 S25942	NADH2 dehydrogenas
28	41	41.8	585	2 T19814	hypothetical prote
29	40.5	41.3	123	2 E71163	hypothetical prote

30	40.5	41.3	318	2 AD2556	hypothetical prote
31	40.5	41.3	666	2 S59109	NADH2 dehydrogenas
32	40.5	41.3	838	2 T47828	hypothetical prote
33	40.5	41.3	1037	2 E84723	hypothetical prote
34	40	40.8	70	2 B82692	hypothetical prote
35	40	40.8	97	2 T03811	Ig heavy chain V r
36	40	40.8	119	2 S36257	hypothetical prote
37	40	40.8	221	2 T28778	ABC-type sugar tra
38	40	40.8	289	2 G96981	probable membrane
39	40	40.8	309	2 AC0489	hypothetical 37.9K
40	40	40.8	337	2 S47742	hypothetical prote
41	40	40.8	337	2 B91179	hypothetical prote
42	40	40.8	337	2 C86025	probable membrane
43	40	40.8	342	2 AE0986	hypothetical prote
44	40	40.8	375	2 C64216	hypothetical prote
45	40	40.8	378	2 G86220	hypothetical prote

ALIGNMENTS

RESULT 1

S61985

ALG6 protein - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein O2513; protein UNA544; protein YOR002w

C;Species: Saccharomyces cerevisiae

C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 20-Jun-2000

C;Accession: S61985; S66867; S72134

R;Sterky, F.; Uhlen, M.

submitted to the EMBL Data Library, December 1995

A;Reference number: S61981

A;Accession: S61985

A;Molecule type: DNA

A;Residues: 1-544 <STE>

A;Cross-references: EMBL:U43491; NID:g1150992; PIDN:AAC49481.1; PID:g1150997

R;Pettersson, B.; Sterky, F.; Uhlen, M.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S66682

A;Accession: S66867

A;Molecule type: DNA

A;Residues: 1-544 <PET>

A;Cross-references: EMBL:Z74910; NID:g1420089; PIDN:CAA99190.1; PID:g1420090; MIPS:

A;Experimental source: strain S288C

R;Sterky, F.; Holmberg, A.; Pettersson, B.; Uhlen, M.

Yeast 12, 1091-1095, 1996

A;Title: The sequence of a 30 kb fragment on the left arm of chromosome XV from Sac

A;Reference number: S72130; MUID:97051599; PMID:8896276

A;Accession: S72134

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-544 <STW>

A;Cross-references: EMBL:U43491; NID:g1150992; PIDN:AAC49481.1; PID:g1150997

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 19

C;Genetics:

A;Gene: SGD:ALG6

A;Cross-references: SGD:S0005528; MIPS:YOR002w

A;Map position: 15R

C;Keywords: transmembrane protein

F;40-56/Domain: transmembrane #status predicted <TM1>

F;145-161/Domain: transmembrane #status predicted <TM2>

F;221-237/Domain: transmembrane #status predicted <TM3>

F;262-278/Domain: transmembrane #status predicted <TM4>

F;336-352/Domain: transmembrane #status predicted <TM5>

F;358-374/Domain: transmembrane #status predicted <TM6>

F;509-525/Domain: transmembrane #status predicted <TM7>

Query Match 51.0%; Score 50; DB 2; Length 544;

Best Local Similarity 61.5%; Pred. No. 3.8;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 NCGIDFIIFWIFW 16

|| : || | |||

Db 514 NCAVGFCFSIFW 526

RESULT 2  
JQ1724  
E1 membrane glycoprotein precursor - canine coronavirus (strain Insavc-1)  
N;Alternate names: matrix glycoprotein  
C;Species: canine coronavirus  
C;Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 16-Jun-2000  
C;Accession: JQ1724  
R;Horsburgh, B.C.; Brierley, I.; Brown, T.D.K.  
J. Gen. Virol. 73, 2849-2862, 1992  
A;Title: Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA  
A;Reference number: PQ0481; MUID:93057357; PMID:1431811  
A;Accession: JQ1724  
A;Molecule type: genomic RNA  
A;Residues: 1-262 <HOR>  
A;Cross-references: DDBJ:D13096; NID:g406193; PIDN:BAA02413.1; PID:g406200  
C;Genetics:  
A;Gene: M  
C;Superfamily: coronavirus E1 membrane glycoprotein  
C;Keywords: glycoprotein; matrix protein; transmembrane protein  
F;1-17/Domain: signal sequence #status predicted <SIG>  
F;18-262/Product: E1 membrane glycoprotein #status predicted <MAT>  
F;56-72/Domain: transmembrane #status predicted <TM1>  
F;85-101/Domain: transmembrane #status predicted <TM2>  
F;115-134/Domain: transmembrane #status predicted <TM3>  
F;32,55/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;33,57/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F;44/Binding site: carbohydrate (Thr) (covalent) #status predicted

Query Match 49.0%; Score 48; DB 1; Length 262;  
Best Local Similarity 50.0%; Pred. No. 4.1;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16  
| | | : | : |  
Db 81 CGIKMLIMWLLW 92

RESULT 3  
D95366  
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) Nuom2 [imported] - Sinorhizobium meliloti  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 03-Jun-2002  
C;Accession: D95366  
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows  
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
A;Reference number: A95262; MUID:21396509; PMID:11481432  
A;Accession: D95366  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-491 <KUR>  
A;Cross-references: GB:AE006469; PIDN:AAK65494.1; PID:g14523966; GSPDB:GN00165  
A;Experimental source: strain 1021, megaplasmid pSymA  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaullt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: nuom2  
A;Genome: plasmid  
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 4  
C;Keywords: oxidoreductase

Query Match 48.0%; Score 47; DB 2; Length 491;  
Best Local Similarity 41.7%; Pred. No. 10;

Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16  
| : | : | : | : |  
Db 127 CALDLFLFYVFW 138

RESULT 4  
A35639  
G protein-coupled receptor RTA - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 16-Jul-1999  
C;Accession: A35639  
R;Ross, P.C.; Figler, R.A.; Corjay, M.H.; Barber, C.M.; Adam, N.; Harcus, D.R.; Ly  
Proc. Natl. Acad. Sci. U.S.A. 87, 3052-3056, 1990  
A;Title: RTA, a candidate G protein-coupled receptor: cloning, sequencing, and tis  
A;Reference number: A35639; MUID:90222168; PMID:2109324  
A;Accession: A35639  
A;Molecule type: mRNA  
A;Residues: 1-343 <ROS>  
A;Cross-references: GB:M35297; NID:g206809; PIDN:AAA42087.1; PID:g206810; GB:M3209  
C;Superfamily: mas transforming protein  
C;Keywords: G protein-coupled receptor; glycoprotein; membrane protein  
F;4/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.9%; Score 45; DB 2; Length 343;  
Best Local Similarity 60.0%; Pred. No. 15;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15  
| | | : | : |  
Db 256 GIDWFLFWVF 265

RESULT 5  
D69890  
conserved hypothetical protein yndM - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: D69890  
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.;  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.;  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulla  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Yosh  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uch  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yosh  
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus sub  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: D69890  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-179 <KUN>  
A;Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13667.1; PID:g26  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: yndM  
C;Superfamily: hypothetical protein ycbP

Query Match 44.9%; Score 44; DB 1; Length 179;  
Best Local Similarity 46.2%; Pred. No. 12;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 ANCGIDFIIFWIF 15  
| : | : | : | : |  
Db 94 ADFGLSFVILWVF 106



RESULT 6  
B96994  
Fe-S oxidoreductase CAC0765 [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: B96994  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: B96994  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-391 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK78741.1; PID:g15023650; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0765

Query Match 44.9%; Score 44; DB 2; Length 391;  
Best Local Similarity 50.0%; Pred. NO. 24;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 ANCGIDFIIFWI 14  
Db 101 SNCGVDFDFYL 112

RESULT 7  
AH0415  
probable membrane-bound sensory transducer in beta-lactamase regulation ampe [imported]  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AH0415  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0415  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-284 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC92652.1; PID:g15981349; GSPDB:GN00175  
C:Genetics:  
A:Gene: ampe

Query Match 43.9%; Score 43; DB 2; Length 284;  
Best Local Similarity 46.7%; Pred. NO. 26;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 QANCGIDFIIFWIFW 16  
Db 184 RAQTGIDGILHLDW 198

RESULT 8  
B81252  
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 Cj1572c [similarity] - Campylobact  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: B81252  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: B81252  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-332 <PAR>  
A:Cross-references: GB:AL139079; GB:AL111168; NID:g6968971; PIDN:CAB73560.1; PID:g6  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: nuoH; Cj1572c  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1  
C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 43.9%; Score 43; DB 2; Length 332;  
Best Local Similarity 75.0%; Pred. NO. 30;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 9 FIIFWIFW 16  
Db 287 FIFFWYFW 294

RESULT 9  
A71937  
amino acid permease - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 20-Aug-1999  
C:Accession: A71937  
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, G  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: A71937  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-475 <ARN>  
A:Cross-references: GB:AE001474; GB:AE001439; NID:g4154929; PIDN:AAD05979.1; PID:g4  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0406  
C:Superfamily: arginine permease

Query Match 43.9%; Score 43; DB 2; Length 475;  
Best Local Similarity 36.4%; Pred. NO. 40;  
Matches 4; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 6 GIDFIIFWIFW 16  
Db 93 GTGYMVFWMYW 103

RESULT 10  
A64647  
amino acid permease (rocE) - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A64647  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann,  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; Mc  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthe  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Frase  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: A64647  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-519 <TOM>  
A:Cross-references: GB:AE000610; GB:AE000511; NID:g2314160; PIDN:AAD08062.1; PID:g2  
C:Superfamily: arginine permease  
C:Keywords: amino acid transport; glycoprotein; transmembrane protein

Query Match 43.9%; Score 43; DB 1; Length 519;  
Best Local Similarity 36.4%; Pred. NO. 43;  
Matches 4; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 6 GIDFIIFWIFW 16  
| : : : : :  
Db 137 GTGYMVFWMYW 147

## RESULT 11

B89823

hypothetical protein arab [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C;Accession: B89823

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura,

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: B89823

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-545 &lt;KUR&gt;

A;Cross-references: GB:BA000018; PID:g13700443; PIDN:BA41741.1; GSPDB:GN00149

A;Experimental source: strain N315

C;Genetics:

A;Gene: arab

C;Superfamily: ribulokinase

Query Match 43.9%; Score 43; DB 2; Length 545;  
Best Local Similarity 50.0%; Pred. No. 45;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 QANCGIDFIIFW 13

| : : : : :  
Db 195 RSNCGLGPKAFW 206

## RESULT 12

YUMU

sucrose synthase (EC 2.4.1.13) - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 11-Jun-1999

C;Accession: S19125

R;Chopra, S.; Del-favero, J.; Dolferus, R.; Jacobs, M.

Plant Mol. Biol. 18, 131-134, 1992

A;Title: Sucrose synthase of Arabidopsis: genomic cloning and sequence characterization.

A;Reference number: S19125; MUID:92119221; PMID:1531031

A;Accession: S19125

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-804 &lt;CHO&gt;

A;Cross-references: EMBL:X60987; NID:g16525; PIDN:CAA43303.1; PID:g16526

C;Genetics:

A;Introns: 31/2; 72/3; 123/2; 187/3; 227/2; 299/3; 331/3; 389/3; 428/3; 484/2; 559/2; 66

C;Superfamily: sucrose synthase; sucrose/sucrose-phosphate synthase homology

C;Keywords: glycosyltransferase; hexosyltransferase

F;276-750/Domain: sucrose/sucrose-phosphate synthase homology &lt;SSPS&gt;

Query Match 43.9%; Score 43; DB 1; Length 804;  
Best Local Similarity 62.5%; Pred. No. 63;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 9 FIIIFWIFW 16

| : : : : :  
Db 19 FLSEFWFVW 26

## RESULT 13

A30560

Ig heavy chain V region (16.4.12E) - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 23-Mar-1989 #sequence\_revision 23-Mar-1989 #text\_change 23-Jul-1999

C;Accession: A30560

R;Matsuda, T.; Kabat, E.A.  
J. Immunol. 142, 863-870, 1989  
A;Title: Variable region cDNA sequences and antigen binding specificity of mouse monoclonal antibody 1A1  
A;Reference number: A30560; MUID:89110062; PMID:2464028  
A;Accession: A30560  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-121 <MAT>  
A;Cross-references: GB:M22955; NID:g195610; PIDN:AAA38368.1; PID:g195611  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin  
F;15-98/Domain: immunoglobulin homology <IMM>

Query Match 42.9%; Score 42; DB 2; Length 121;  
Best Local Similarity 50.0%; Pred. No. 18;  
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 ANCGIDFIIFWIFW 16

| : : : : :  
Db 23 AASGIDFSRYWMSW 36

## RESULT 14

S64252

probable membrane protein YGL230c - yeast (Saccharomyces cerevisiae)

N;Alternate names: hypothetical protein G0946

C;Species: Saccharomyces cerevisiae

C;Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 19-Apr-2002

C;Accession: S64252

R;Fartmann, B.; Kramer, B.; Kramer, W.

submitted to the Protein Sequence Database, May 1996

A;Reference number: S64248

A;Accession: S64252

A;Molecule type: DNA

A;Residues: 1-147 &lt;FAR&gt;

A;Cross-references: EMBL:Z72752; NID:g1322886; PID:g1322887; GSPDB:GN00007; MIPS:YGL230c

A;Experimental source: strain S288C

C;Genetics:

A;Gene: MIPS:YGL230c

A;Cross-references: SGD:S0003199

A;Map position: 7L

C;Superfamily: Saccharomyces cerevisiae probable membrane protein YGL230c

C;Keywords: transmembrane protein

F;89-105/Domain: transmembrane #status predicted &lt;TMM&gt;

Query Match 42.9%; Score 42; DB 2; Length 147;  
Best Local Similarity 50.0%; Pred. No. 21;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16

| : : : : :  
Db 96 CAIQFLFFIYW 107

## RESULT 15

AB2320

hypothetical protein all4113 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp.

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C;Accession: AB2320

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri-

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tab-

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Nostoc sp. strain PCC 7120

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AB2320

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-194 &lt;KUR&gt;

A;Cross-references: GB:BA000019; PIDN:BA75812.1; PID:g17133248; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:  
A:Gene: all4113  
C;Superfamily: Synechocystis hypothetical protein sll1656  
Query Match 42.9%; Score 42; DB 2; Length 194;  
Best Local Similarity 37.5%; Pred. No. 27;  
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
QY 1 FQANCGIDFIIFWIFW 16  
| | | | : | : |  
Db 70 FWA VCGVIALLF SVY W 85

Search completed: July 9, 2003, 15:09:49  
Job time : 12.2212 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 1.9823 Seconds  
(without alignments)  
334.773 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	51.0	544	1	ALG6_YEAST
2	48	49.0	262	1	VME1_CVCAI
3	45	45.9	343	1	RTA_RAT
4	45	45.9	406	1	G6AB_DROME
5	43	43.9	545	1	ARAB_STAAM
6	43	43.9	804	1	SUS2_ARATH
7	42	42.9	147	1	YGY0_YEAST
8	42	42.9	329	1	YL3L_ARCFU
9	42	42.9	389	1	FUT7_MOUSE
10	41	41.8	342	1	Y762_METJA
11	41	41.8	495	1	NU4M_MARPO
12	40.5	41.3	666	1	NU5M_CHOCR
13	40	40.8	337	1	YHJD_ECOLI
14	40	40.8	375	1	Y147_MYCGE
15	40	40.8	593	1	DCEL_RAT
16	40	40.8	594	1	DCEL_FELCA
17	40	40.8	594	1	DCEL_HUMAN
18	40	40.8	594	1	DCEL_PIG
19	39.5	40.3	107	1	ELBA_ECOLI
20	39	39.8	262	1	VGLL_CVPR8
21	39	39.8	262	1	VME1_CVPFS
22	39	39.8	262	1	VME1_CVPPU
23	39	39.8	262	1	VME1_CVPRM
24	39	39.8	262	1	VME1_FIPV
25	39	39.8	310	1	RCEL_CHLAU
26	39	39.8	329	1	MAY_LACIA
27	39	39.8	440	1	YA85_MYCPN
28	39	39.8	679	1	NADE_TREPA
29	38	38.8	116	1	HV36_MOUSE
30	38	38.8	117	1	HV42_MOUSE
31	38	38.8	118	1	HV39_MOUSE
32	38	38.8	119	1	HV37_MOUSE
33	38	38.8	119	1	HV38_MOUSE

34	38	38.8	119	1	HV40_MOUSE	P01810	mus musculus
35	38	38.8	236	1	CD63_BOVIN	Q9xsk2	bos taurus
36	38	38.8	245	1	TATC_CAMJE	Q9pht8	campylobacter
37	38	38.8	282	1	AQP6_HUMAN	Q13520	homo sapiens
38	38	38.8	283	1	NADC_METJA	Q57916	methanococcus
39	38	38.8	317	1	YKG8_CAEEL	P46558	caenorhabditis
40	38	38.8	325	1	MC5R_MOUSE	P41149	mus musculus
41	38	38.8	325	1	MC5R_RAT	P35345	rattus norvegicus
42	38	38.8	336	1	MRAY_STRPY	Q99yk2	streptococcus
43	38	38.8	377	1	Y147_MYCPN	P75585	mycoplasma
44	38	38.8	398	1	Y432_MYCGE	Q49432	mycoplasma
45	38	38.8	404	1	Y432_MYCPN	P75166	mycoplasma

ALIGNMENTS

RESULT 1  
ALG6\_YEAST  
ID ALG6\_YEAST STANDARD; PRT; 544 AA.  
AC Q12001;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase (EC 2.4.1.-) (Dolichyl-P-Glc:Man9GlcNAc2-PP-dolichyl glucosyltransferase).  
GN ALG6 OR YOR002W OR UNA544.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97051599; PubMed=8896276;  
RA Sterky F., Holmberg A., Pettersson B., Uhlen M.;  
RT "The sequence of a 30 kb fragment on the left arm of chromosome XV from Saccharomyces cerevisiae reveals 15 open reading frames, five of which correspond to previously identified genes.";  
RL Yeast 12:1091-1095(1996).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=97031444; PubMed=8877369;  
RA Reiss G., te Heesen S., Zimmerman J., Robbins P.W., Aebi M.;  
RT "Isolation of the ALG6 locus of Saccharomyces cerevisiae required for glucosylation in the N-linked glycosylation pathway.";  
RL Glycobiology 6:493-498(1996).  
CC -!- FUNCTION: Adds the first glucose residue to the lipid-linked oligosaccharide precursor for N-linked glycosylation. Transfers glucose from dolichyl phosphate glucose (Dol-P-Glc) onto the lipid-linked oligosaccharide Man(9)GlcNAc(2)-pp-Dol.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (Potential).  
CC -!- SIMILARITY: BELONGS TO THE ALG6/ALG8 GLUCOSYLTRANSFERASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U43491; AAC49481.1; -.  
DR EMBL; Z74910; CAA99190.1; -.  
DR SGD; S0005528; ALG6.  
DR InterPro; IPR004856; Alg6\_Alg8.  
DR Pfam; PF03155; Alg6\_Alg8; 1.  
KW Transferase; Glycosyltransferase; Transmembrane;  
KW Endoplasmic reticulum.  
FT TRANSMEM 36 56 POTENTIAL.  
FT TRANSMEM 105 125 POTENTIAL.  
FT TRANSMEM 146 166 POTENTIAL.



```
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 224 244 POTENTIAL.
FT TRANSMEM 264 284 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 426 446 POTENTIAL.
FT TRANSMEM 482 502 POTENTIAL.
FT TRANSMEM 509 529 POTENTIAL.
SQ SEQUENCE 544 AA; 62782 MW; 64BFA11A1F6D02B7 CRC64;

Query Match 51.0%; Score 50; DB 1; Length 544;
Best Local Similarity 61.5%; Pred. No. 1.3;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 4 NCGIDFIIFWIFW 16
||| :| | |
Db 514 NCAVGFICFSIFW 526

RESULT 2
VME1_CVCAI STANDARD; PRT; 262 AA.
AC P36299;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE E1 glycoprotein precursor (Matrix glycoprotein) (Membrane
glycoprotein).
GN M.
OS Canine enteric coronavirus (strain Insavc-1) (CCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=36391;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93057357; PubMed=1431811;
RA Horsburgh B.C., Brierley I., Brown T.D.K.;
RT "Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus
genomic RNA.";
RL J. Gen. Virol. 73:2849-2862(1992).
CC -!- FUNCTION: TARGETS THE SITE OF VIRUS MORPHOGENESIS AND MAY BE
IMPLICATED IN VIRAL PATHOGENESIS.
CC -!- SUBCELLULAR LOCATION: LARGELY EMBEDDED IN THE LIPID BILAYER.
CC -!- SIMILARITY: BELONGS TO THE CORONAVIRUS M PROTEIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D13096; BAA02413.1; -.
DR PIR; JQ1724; JQ1724.
DR InterPro; IPR002574; Corona_M.
DR Pfam; PF01635; Corona_M; 1.
KW Matrix protein; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 262 E1 GLYCOPROTEIN.
FT TRANSMEM 56 72 POTENTIAL.
FT TRANSMEM 85 101 POTENTIAL.
FT TRANSMEM 115 134 POTENTIAL.
FT CARBOHYD 32 32 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 33 33 O-LINKED (POTENTIAL).
FT CARBOHYD 44 44 O-LINKED (POTENTIAL).
FT CARBOHYD 55 55 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 262 AA; 29508 MW; 35C6D45DC6A474F6 CRC64;

Query Match 49.0%; Score 48; DB 1; Length 262;
Best Local Similarity 50.0%; Pred. No. 1.4;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 5 CGIDFIIFWIFW 16
||| :| | |
Db 81 CGIKMLIMWLLW 92

RESULT 3
RTA_RAT STANDARD; PRT; 343 AA.
ID RTA_RAT
AC P23749;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Probable G protein-coupled receptor RTA.
GN RTA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Aorta;
RX MEDLINE=90222168; PubMed=2109324;
RA Ross P.C., Figler R.A., Corjay M.H., Barber C.M., Adam N.,
Harcus D.R., Lynch K.R.;
RT "RTA, a candidate G protein-coupled receptor: cloning, sequencing,
and tissue distribution.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:3052-3056(1990).
CC -!- FUNCTION: ORPHAN RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: GUT, VAS DEFERENS, UTERUS, AND AORTA BUT ONLY
BARELY DETECTABLE IN LIVER, KIDNEY, LUNG, AND SALIVARY GLAND. IN
THE BRAIN, RTA IS MARKEDLY ABUNDANT IN THE CEREBELLUM.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
MOST SIMILAR TO MAS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M35297; AAA42087.1; -.
DR EMBL; M35298; AAA42088.1; -.
DR PIR; A35639; A35639.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 44 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 45 66 1 (POTENTIAL).
FT DOMAIN 67 82 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 83 104 2 (POTENTIAL).
FT DOMAIN 105 123 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 124 144 3 (POTENTIAL).
FT DOMAIN 145 160 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 161 181 4 (POTENTIAL).
FT DOMAIN 182 198 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 199 220 5 (POTENTIAL).
FT DOMAIN 221 241 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 242 263 6 (POTENTIAL).
FT DOMAIN 264 273 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 274 294 7 (POTENTIAL).
FT DOMAIN 295 343 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 4 4 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 343 AA; 38364 MW; E4630007770941F4 CRC64;

Query Match 45.9%; Score 45; DB 1; Length 343;
Best Local Similarity 60.0%; Pred. No. 5.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      6 GIDFIIFWIF 15
Db      256 GIDWFLFWVF 265

RESULT 4
G64B_DROME      STANDARD;      PRT;      406 AA.
AC  P83294; Q9VZJ7;
DT  15-JUN-2002 (Rel. 41, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Putative gustatory receptor 64b.
GN  GR64B OR CG14986.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC  Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Berkley;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Ballev R.M., Basu A., Baxendale J., Andrews-Pfannkuch C., Baldwin D.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA  Burtis J.M., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry K.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA  Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
RN  [2]
RP  IDENTIFICATION.
RX  MEDLINE=21407712; PubMed=11516643;
RA  Dunipace L., Meister S., McNeely C., Amrein H.;
RT  "Spatially restricted expression of candidate taste receptors in the
RT  Drosophila gustatory system.";
RL  Curr. Biol. 11:822-835(2001).
RN  [3]
RP  CONCEPTUAL TRANSLATION.
RA  Robertson H.;
RL  Unpublished observations (NOV-2001).
```

```
CC  FUNCTION: Probable role in the gustatory response.
CC  SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  SIMILARITY: BELONGS TO FAMILY DR-TR OF G-PROTEIN COUPLED
CC  RECEPTORS, SUBFAMILY II.
CC  CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC  gene model prediction.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE003480; AAF47824.1; ALT_SEQ.
DR  FlyBase; FBgn0045478; Gr64b.
KW  Hypothetical protein; Receptor; G-protein coupled receptor;
KW  Transmembrane; Glycoprotein; Multigene family.
FT  DOMAIN 1 47 EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM 48 68 1 (POTENTIAL).
FT  DOMAIN 69 79 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 80 100 2 (POTENTIAL).
FT  DOMAIN 101 130 EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM 131 151 3 (POTENTIAL).
FT  DOMAIN 152 183 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 184 204 4 (POTENTIAL).
FT  DOMAIN 205 265 EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM 266 286 5 (POTENTIAL).
FT  DOMAIN 287 290 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 291 311 6 (POTENTIAL).
FT  DOMAIN 312 370 EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM 371 391 7 (POTENTIAL).
FT  DOMAIN 392 406 CYTOPLASMIC (POTENTIAL).
FT  CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE 406 AA; 47488 MW; F6A37DD9E1E58B88 CRC64;

Query Match 45.9%; Score 45; DB 1; Length 406;
Best Local Similarity 46.2%; Pred. No. 6;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      4 NCIGDPIIFWIFW 16
Db      291 NIGVDFLVLAFW 303

RESULT 5
ARAB_STAAM      STANDARD;      PRT;      545 AA.
AC  Q99W57;
DT  15-JUN-2002 (Rel. 41, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  L-ribulokinase (EC 2.7.1.16).
GN  ARAB OR SAV0552 OR SA0510.
OS  Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS  Staphylococcus aureus (strain N315).
OC  Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX  NCBI_TaxID=158878, 158879;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Mu50 / ATCC 700699, and N315;
RX  MEDLINE=21311952; PubMed=11418146;
RA  Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA  Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA  Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA  Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA  Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA  Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA  Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT  "Whole genome sequencing of methicillin-resistant Staphylococcus
RT  aureus.";
RL  Lancet 357:1225-1240(2001).
```

CC -!- CATALYTIC ACTIVITY: ATP + L-ribose = ADP + L-ribose 5-phosphate.  
CC -!- PATHWAY: L-arabinose catabolism; second step.  
CC -!- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AP003359; BAB56714.1; -.  
DR EMBL; AP003130; BAB41741.1; -.  
DR InterPro; IPR000577; FGGY\_kin.  
DR Pfam; PF00370; FGGY; 1.  
DR Pfam; PF02782; FGGY\_C; 1.  
KW Transferase; Kinase; Arabinose catabolism; Complete proteome.  
SQ SEQUENCE 545 AA; 60965 MW; E3917036237CEA4B CRC64;  
  
Query Match 43.9%; Score 43; DB 1; Length 545;  
Best Local Similarity 50.0%; Pred. No. 16;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 QANCGIDFIIFW 13  
Db 195 RSNCGLGFKAFW 206  
::|||: | ||  
  
RESULT 6  
SUS2\_ARATH STANDARD; PRT; 804 AA.  
AC Q00917;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Sucrose synthase (EC 2.4.1.13) (Sucrose-UDP glucosyltransferase).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=cv. Columbia;  
RX MEDLINE=92119221; PubMed=1531031;  
RA Chopra S., Del-Favero J., Dolferus R., Jacobs M.;  
RT "Sucrose synthase of Arabidopsis: genomic cloning and sequence  
RT characterization."  
RL Plant Mol. Biol. 18:131-134 (1992).  
CC -!- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND  
CC FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.  
CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose = UDP + sucrose.  
CC -!- INDUCTION: BY ANAEROBIC STRESS.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1. PLANT  
CC SUCROSE SYNTHASE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X60987; CAA43303.1; -.  
DR PIR; S19125; YUMU.  
DR InterPro; IPR001296; Glycos\_transf\_1.  
DR InterPro; IPR000368; Sucrose\_synth.  
DR Pfam; PF00534; Glycos\_transf\_1; 1.  
DR Pfam; PF00862; Sucrose\_synth; 1.  
KW Transferase; Glycosyltransferase; Multigene family.

SQ SEQUENCE 804 AA; 91989 MW; 3E727D3CDFF9A4B9 CRC64;  
  
Query Match 43.9%; Score 43; DB 1; Length 804;  
Best Local Similarity 62.5%; Pred. No. 23;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 9 FIIFWIFW 16  
Db 19 FLSFWVFW 26  
|: ||:|  
  
RESULT 7  
YGY0\_YEAST STANDARD; PRT; 147 AA.  
AC P53074;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Hypothetical 17.3 kDa protein in SEC15-SAP4 intergenic region.  
GN YGL230C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fartmann B., Kramer B., Kramer W.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 272752; CAA96948.1; -.  
DR SGD; S0003199; YGL230C.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 85 105 POTENTIAL.  
SQ SEQUENCE 147 AA; 17262 MW; F8F040A1D618CD96 CRC64;  
  
Query Match 42.9%; Score 42; DB 1; Length 147;  
Best Local Similarity 50.0%; Pred. No. 7.2;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 5 CGIDFIIFWIFW 16  
Db 96 CAIQFLFFIIYW 107  
| | | : | | |  
  
RESULT 8  
YL31\_ARCFU STANDARD; PRT; 329 AA.  
AC O28149;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AF2131.  
GN AF2131.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,



RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artiaich P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000957; AAB89128.1; -.  
CC TIGR; AF2131; -.  
DR DR

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 5 24 POTENTIAL.  
FT TRANSMEM 34 56 POTENTIAL.  
FT TRANSMEM 92 114 POTENTIAL.  
FT TRANSMEM 124 146 POTENTIAL.  
FT TRANSMEM 159 181 POTENTIAL.  
FT TRANSMEM 196 218 POTENTIAL.  
FT TRANSMEM 231 253 POTENTIAL.  
FT TRANSMEM 263 285 POTENTIAL.  
FT TRANSMEM 306 328 POTENTIAL.  
SQ SEQUENCE 329 AA; 37066 MW; C07CAF427DC26F7 CRC64;

Query Match 42.9%; Score 42; DB 1; Length 329;  
Best Local Similarity 71.4%; Pred. No. 15;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 10 IIFWIFW 16  
I::I::I  
Db 304 ILFWLFW 310

## RESULT 9

FUT7\_MOUSE  
ID FUT7\_MOUSE STANDARD; PRT; 389 AA.  
AC Q11131;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Alpha-(1,3)-fucosyltransferase (EC 2.4.1.-) (Galactoside 3-L-  
DE fucosyltransferase) (Fucosyltransferase 7) (FUCT-VII).  
GN FUT7.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIH Swiss;  
RX MEDLINE=96215226; PubMed=8626519;  
RA Smith P.L., Gersten K.M., Petryniak B., Kelly R.J., Rogers C.,  
RA Natsuka Y., Alford J.A. III, Scheidegger E.P., Natsuka S., Lowe J.B.;  
RT "Expression of the alpha(1,3)fucosyltransferase Fuc-TVII in lymphoid  
RT aggregate high endothelial venules correlates with expression of L-  
RT selectin ligands.";  
RL J. Biol. Chem. 271:8250-8259(1996).  
CC -!- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN  
CC THE EXPRESSION OF SIALYL LEWIS X ANTIGENS.  
CC -!- CATALYTIC ACTIVITY: GDP-L-fucose + alpha-2,3-Neu-N-acetyl-1,4-  
CC beta-D-galactosyl-N-acetyl-D-glucosaminyl-R = GDP + alpha-2,3-  
CC Neu-N-acetyl-1,4-beta-D-galactosyl-(alpha-1,3-L-fucosyl)-  
CC N-acetyl-D-glucosaminyl-R.

CC -!- PATHWAY: Glycosylation.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND  
CC FORM IN TRANS CISTERNAE OF GOLGI.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG AND BONE MARROW AND  
CC TO A MUCH LESSER EXTENT IN SPLEEN, SALIVARY GLAND AND SKELETAL  
CC MUSCLE.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 10.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U45980; AAC52484.1; -.  
CC EMBL; U45980; AAC52485.1; -.  
CC MGD; MGI:107692; Fut7.  
DR InterPro; IPR001503; GT\_10.  
DR Pfam; PF00852; Glyco\_transf\_10; 1.  
KW Transferase; Glycosyltransferase; Transmembrane; Glycoprotein;  
KW Signal-anchor; Golgi stack; Alternative splicing.  
FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 56 78 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT DOMAIN 79 389 LUMENAL, CATALYTIC (POTENTIAL).  
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 1 51 MPTPCPPACLSPTGTHRLLPFDWKAPSWESKKEATCNSSS  
FT PGPWAEPTVQ -> MNCI (IN ISOFORM 2).  
SQ SEQUENCE 389 AA; 44494 MW; 118FC6B2378B99C6 CRC64;

Query Match 42.9%; Score 42; DB 1; Length 389;  
Best Local Similarity 45.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 GIDFIIFWIFW 16  
I::I::I  
Db 68 GATFMVIVFW 78

## RESULT 10

Y762\_METJA  
ID Y762\_METJA STANDARD; PRT; 342 AA.  
AC Q58172;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MJ0762.  
GN MJ0762.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

```
CC -!- SIMILARITY: STRONG, TO M.JANNASCHII MJ0576 AND TO S.POMBE MALATE
CC PERMEASE (MAE1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67521; AAB98753.1; -.
CC TIGR; MJ0762; -.
CC InterPro; IPR004695; C4dic_mal_transp.
CC Pfam; PF03595; C4dic_mal_tran; 1.
CC TIGRFAMs; TIGR00816; tdt; 1.
CC Hypothetical protein; Transmembrane; Transport; Complete proteome.
CC TRANSMEM 8 28 POTENTIAL.
CC TRANSMEM 39 59 POTENTIAL.
CC TRANSMEM 79 99 POTENTIAL.
CC TRANSMEM 108 128 POTENTIAL.
CC TRANSMEM 142 162 POTENTIAL.
CC TRANSMEM 175 195 POTENTIAL.
CC TRANSMEM 207 227 POTENTIAL.
CC TRANSMEM 242 262 POTENTIAL.
CC TRANSMEM 276 296 POTENTIAL.
CC TRANSMEM 304 324 POTENTIAL.
CC SEQUENCE 342 AA; 39534 MW; 08EFEC3E2C4955D8 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 342;
Best Local Similarity 45.0%; Pred. No. 22;
Matches 9; Conservative 1; Mismatches 0; Indels 10; Gaps 1;

QY 6 GIDFI-----IFWIF 15
    |||||
    |||||
Db 93 GIDFILIKNNLFLGKIFWVF 112

RESULT 11
NU4M_MARPO
ID NU4M_MARPO STANDARD; PRT; 495 AA.
AC P26848;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
GN ND4 OR NAD4.
OS Marchantia polymorpha (Liverwort).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114051; PubMed=1731062;
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;
RT "Gene organization deduced from the complete sequence of liverwort
RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant
RT mitochondrial genome.";
RL J. Mol. Biol. 223:1-7(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93247547; PubMed=8483448;
RA Nozato N., Oda K., Yamato K., Ohta E., Takemura M., Akashi K.,
RA Fukuzawa H., Ohyama K.;
RT "Cotranscriptional expression of mitochondrial genes for subunits of
RT NADH dehydrogenase, nad5, nad4, nad2, in Marchantia polymorpha.";
RL Mol. Genet. 237:343-350(1993).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M68929; AAC09398.1; -.
CC PIR; S25942; S25942.
CC InterPro; IPR003918; NADHub_oxred4.
CC InterPro; IPR001750; Oxidored_q1.
CC Pfam; PF00361; oxidored_q1; 1.
CC PRINTS; PR01437; NUOXDRDTASE4.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC SEQUENCE 495 AA; 56311 MW; 0F7589AD6CAAAED4 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 495;
Best Local Similarity 41.7%; Pred. No. 31;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 CGIDFIIFWIFW 16
    | : | : | : | :
Db 135 CSLDLLIFVVF 146

RESULT 12
NU5M_CHOCH
ID NU5M_CHOCH STANDARD; PRT; 666 AA.
AC P48920;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN ND5 OR NAD5.
OS Chondrus crispus (Carrageen).
OG Mitochondrion.
OC Eukaryota; Rhodophyta; Florideophyceae; Gigartinales; Gigartineae;
OC Chondrus.
OX NCBI_TaxID=2769;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Apices;
RX MEDLINE=95341681; PubMed=7616569;
RA Leblanc C., Boyen C., Richard O., Bonnard G., Grienberger J.M.,
RA Kloeareg B.;
RT "Complete sequence of the mitochondrial DNA of the rhodophyte
RT Chondrus crispus (Gigartinales). Gene content and genome
RT organization.";
RL J. Mol. Biol. 250:484-495(1995).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z47547; CAA87625.1; -.
CC InterPro; IPR003916; NADHub_oxred5.
CC InterPro; IPR001750; Oxidored_q1.
CC InterPro; IPR001516; Oxidored_q1_N.
CC Pfam; PF00361; oxidored_q1; 1.
CC Pfam; PF00662; oxidored_q1_N; 1.
CC PRINTS; PR01434; NADHDHGNASE5.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC SEQUENCE 666 AA; 75628 MW; 37E86F2C24B9D360 CRC64;

Query Match 41.3%; Score 40.5; DB 1; Length 666;
Best Local Similarity 53.3%; Pred. No. 48;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;
```



QY 1 FOANCGIDFIIFWIF 15  
I : |||:||||  
Db 176 FGLSLGI-FLIFWVF 189

## RESULT 13

YHJD\_ECOLI  
ID YHJD\_ECOLI STANDARD; PRT; 337 AA.

AC P37642;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein yhjD.

GN YHJD OR B3522.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=K12 / MG1655;

RX MEDLINE=94316500; PubMed=8041620;

RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;

RT "Analysis of the Escherichia coli genome. V. DNA sequence of the

RT region from 76.0 to 81.5 minutes.";

RL Nucleic Acids Res. 22:2576-2586(1994).

CC -!- SIMILARITY: STRONG, TO E.CHRYSANTHEMI HYPOTHETICAL PROTEIN IN

CC KDGK 5'REGION (AC P45417).

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U00039; AAB18498.1; -.

DR EMBL; AE000428; AAC76547.1; -.

DR EcoGene; EG12248; yhjD.

DR InterPro; IPR005274; Cons\_hypoth766.

DR InterPro; IPR004664; RNase\_BN.

DR Pfam; PF03631; Ribonuclease\_BN; 1.

DR TIGRFAMs; TIGR00766; TIGR00766; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 337 AA; 37911 MW; C41B2A224902E311 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 337;

Best Local Similarity 71.4%; Pred. No. 31;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 10 IIFWIFW 16

: |||||

Db 244 LFFWIFW 250

## RESULT 14

Y147\_MYCGE

ID Y147\_MYCGE STANDARD; PRT; 375 AA.

AC P47393;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MG147.

GN MG147.

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,

RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U39695; AAC71365.1; -.

DR TIGR; MG147; -.

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 21 41 POTENTIAL.

FT TRANSMEM 66 86 POTENTIAL.

FT TRANSMEM 161 181 POTENTIAL.

FT TRANSMEM 203 223 POTENTIAL.

FT TRANSMEM 234 254 POTENTIAL.

FT TRANSMEM 289 309 POTENTIAL.

FT TRANSMEM 338 358 POTENTIAL.

SQ SEQUENCE 375 AA; 43188 MW; A14AF07D574E8046 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 375;

Best Local Similarity 55.6%; Pred. No. 34;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 8 DFIFWIFW 16

|||: |||:

Db 239 DFLVLWIFY 247

## RESULT 15

DCE1\_RAT

ID DCE1\_RAT STANDARD; PRT; 593 AA.

AC P18088;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Glutamate decarboxylase, 67 kDa isoform (EC 4.1.1.15) (GAD-67)

DE (67 kDa glutamic acid decarboxylase).

GN GAD1 OR GAD67.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91014554; PubMed=2170798;

RA Wyborski R.J., Bond R.W., Gottlieb D.I.;

RT "Characterization of a cDNA coding for rat glutamic acid

RT decarboxylase.";

RL Brain Res. Mol. Brain Res. 8:193-198(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=90132703; PubMed=2299361;

RA Julien J.F., Samama P., Mallet J.;

RT "Rat brain glutamic acid decarboxylase sequence deduced from a cloned

RT cDNA.";

RL J. Neurochem. 54:703-705(1990).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=92020930; PubMed=1924335;

RA Michelsen B.K., Petersen J.S., Boel E., Moldrup A., Dyrberg T.,

RA Madsen O.D.;

RT "Cloning, characterization, and autoimmune recognition of rat islet

RT glutamic acid decarboxylase in insulin-dependent diabetes mellitus.";

```
RL Proc. Natl. Acad. Sci. U.S.A. 88:8754-8758(1991).
CC -!- FUNCTION: CATALYZES THE PRODUCTION OF GABA.
CC -!- CATALYTIC ACTIVITY: L-glutamate = 4-aminobutanoate + CO(2).
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
CC TYRDC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M34445; AAC42037.1; -.
DR EMBL; X57572; CAA40800.1; -.
DR EMBL; X57573; CAA40801.1; -.
DR EMBL; M76177; AAA41184.1; -.
DR PIR; A41367; A41367.
DR PIR; A43756; A43756.
DR InterPro; IPR002129; Pyridoxal_dec.
DR Pfam; PF00282; pyridoxal_dec; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Neurotransmitter biosynthesis; Lyase; Decarboxylase;
KW Pyridoxal phosphate; Multigene family.
FT BINDING 404 404 PYRIDOXAL PHOSPHATE (POTENTIAL).
FT CONFLICT 103 103 L -> V (IN REF. 2).
FT CONFLICT 284 284 F -> S (IN REF. 2).
FT CONFLICT 287 288 EH -> AD (IN REF. 2).
FT CONFLICT 344 345 AG -> EA (IN REF. 2).
FT CONFLICT 347 347 T -> I (IN REF. 2).
FT CONFLICT 352 353 FD -> LE (IN REF. 2).
FT CONFLICT 380 380 L -> R (IN REF. 2).
SQ SEQUENCE 593 AA; 66640 MW; EF83239C30301F69 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 593;
Best Local Similarity 42.9%; Pred. No. 52;
Matches 6; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

QY 5 CG--IDFIIFWIFW 16
|| :| ||: |
Db 454 CGRHVDIFKFWLMW 467

Search completed: July 9, 2003, 15:08:58
Job time : 11.9823 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 15:05:28 ; Search time 7.22124 Seconds  
(without alignments)  
456.536 Million cell updates/sec

Title: US-09-854-133-587  
Perfect score: 98  
Sequence: 1 FQANCGIDFIIFWIFW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	50.0	495	1 Q9P9F4	Q9p9f4 methanosarc
2	49	50.0	495	17 Q8TQN5	Q8tqn5 methanosarc
3	47	48.0	275	8 Q9TDU3	Q9tdu3 echinococcu
4	47	48.0	491	2 Q9EV68	Q9ev68 rhizobium m
5	47	48.0	491	16 Q92YNG	Q92yn6 rhizobium m
6	47	48.0	535	8 Q958U2	Q958u2 echinococcu
7	46	46.9	192	5 Q8T4L1	Q8t4l1 brugia mala
8	46	46.9	310	11 Q8VEV1	Q8vev1 mus musculu
9	46	46.9	530	8 Q953M3	Q953m3 echinococcu
10	46	46.9	3223	2 Q9RPH1	Q9rph1 escherichia
11	46	46.9	3223	2 Q9RM48	Q9rm48 escherichia
12	46	46.9	3223	2 Q8VQR2	Q8vqr2 escherichia
13	46	46.9	3223	2 Q8VNR6	Q8vnr6 escherichia
14	45	45.9	269	5 Q968Z8	Q968z8 plasmodium
15	45	45.9	319	11 Q91ZB6	Q91zb6 mus musculu
16	45	45.9	343	4 Q96AM1	Q96am1 homo sapien

17	45	45.9	343	11	Q8VVCJ6	Q8vcj6 mus musculu
18	45	45.9	856	5	Q9NA34	Q9na34 caenorhabdi
19	44	44.9	149	8	Q02708	Q02708 podospora a
20	44	44.9	179	16	Q31816	Q31816 bacillus su
21	44	44.9	391	16	Q97L01	Q97l01 clostridium
22	44	44.9	531	5	Q9U8R4	Q9u8r4 drosophila
23	44	44.9	899	5	Q95XS9	Q95xs9 caenorhabdi
24	44	44.9	1440	5	Q9V3S5	Q9v3s5 drosophila
25	43	43.9	167	4	Q96DC5	Q96dc5 homo sapien
26	43	43.9	179	4	Q9UHQ2	Q9uhq2 homo sapien
27	43	43.9	221	4	Q9UHQ3	Q9uhq3 homo sapien
28	43	43.9	221	4	Q9BRV3	Q9brv3 homo sapien
29	43	43.9	221	6	Q95KW8	Q95kw8 papio anubi
30	43	43.9	284	16	Q8ZBI8	Q8zbi8 yersinia pe
31	43	43.9	332	16	Q9PMA3	Q9pma3 campylobact
32	43	43.9	475	16	Q9ZM17	Q9zml7 helicobacte
33	43	43.9	503	17	Q9HJK7	Q9hjk7 thermoplas
34	43	43.9	516	17	Q979N1	Q979n1 thermoplas
35	43	43.9	519	16	Q25661	Q25661 helicobacte
36	43	43.9	530	8	Q9B8V3	Q9b8v3 schistosoma
37	43	43.9	805	10	Q9FJ20	Q9fj20 arabidopsis
38	42	42.9	194	16	Q8YPS8	Q8yps8 anabaena sp
39	42	42.9	256	10	Q9SB30	Q9sb30 arabidopsis
40	42	42.9	351	5	Q76204	Q76204 schistosoma
41	42	42.9	386	3	Q93875	Q93875 candida alb
42	42	42.9	483	8	Q36097	Q36097 theileria p
43	42	42.9	622	13	Q57661	Q57661 xenopus lae
44	42	42.9	671	10	Q9FMN5	Q9fmn5 arabidopsis
45	41	41.8	103	8	Q9G1S1	Q9g1s1 venerupis p

ALIGNMENTS

RESULT 1  
Q9P9F4  
ID Q9P9F4 PRELIMINARY; PRT; 495 AA.  
AC Q9P9F4;  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)  
DE F420H2 dehydrogenase subunit FpOM.  
GN FPOM.  
OS Methanosarcina mazei (Methanosarcina frisia).  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2209;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G01;  
RA Abken H.-J., Deppenmeier U.;  
RT "Purification and properties of an F420H2 dehydrogenase from Methanosarcina mazei Gol.";  
RL FEMS Microbiol. Lett. 154:231-237(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G01;  
RX MEDLINE=20309738; PubMed=10751389;  
RA Baumer S., Ide T., Jacobi C., Johann A., Gottschalk G.,  
RA Deppenmeier U.;  
RT "The F420H2 Dehydrogenase from Methanosarcina mazei Is a Redox-driven Proton Pump Closely Related to NADH Dehydrogenases.";  
RL J. Biol. Chem. 275:17968-17973(2000).  
DR EMBL; AF228525; AAF65740.1; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR001750; Oxidored\_q1.  
DR Pfam; PF00361; oxidored\_q1; 1.  
DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
KW NAD; Oxidoreductase.  
SQ SEQUENCE 495 AA; 53972 MW; F288A37630CAFA00 CRC64;  
Query Match 50.0%; Score 49; DB 1; Length 495;  
Best Local Similarity 60.0%; Pred. No. 7.9;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 7 IDFIIFWIFW 16  
:|:|:|:|:|  
Db 128 LDFVVFYIFW 137

## RESULT 2

Q8TQN5  
ID Q8TQN5 PRELIMINARY; PRT; 495 AA.  
AC Q8TQN5;  
DT 01-JUN-2002 (TReMBLrel. 21, Created)  
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE F(420)H(2) dehydrogenase, subunit FpOM.  
GN FpOM OR MA1505.  
OS Methanosarcina acetivorans.  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2214;  
RN [1]  
RP SEQUENCE FROM N.A. 35395 / DSM 2834;  
RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
RX MEDLINE=21929760; PubMed=11932238;  
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,  
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,  
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,  
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,  
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,  
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
RA Metcalf W.W., Birren B.;  
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
RT and physiological diversity."  
RL Genome Res. 12:532-542(2002).  
DR EMBL; AE010819; AAM04919.1; -.  
KW Complete proteome.  
SQ SEQUENCE 495 AA; 54138 MW; 278918B16BA7BAE6 CRC64;

Query Match 50.0%; Score 49; DB 17; Length 495;  
Best Local Similarity 60.0%; Pred. No. 7.9;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 7 IDFIIFWIFW 16  
:|:|:|:|:|  
Db 128 LDFVVFYIFW 137

## RESULT 3

Q9TDU3  
ID Q9TDU3 PRELIMINARY; PRT; 275 AA.  
AC Q9TDU3;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).  
GN COL.  
OS Echinococcus granulosus.  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;  
OC Cyclophyllidae; Taeniidae; Echinococcus.  
OX NCBI\_TaxID=6210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21036605; PubMed=11163447;  
RA Nakao M., Sako Y., Yokoyama N., Fukunaga M., Ito A.;  
RT "Mitochondrial genetic code in cestodes."  
RL Mol. Biochem. Parasitol. 111:415-424(2000).  
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-

CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 4 FERRICYTOCHROME  
CC C + 2 H(2)O.  
CC -!- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
DR EMBL; AB033407; BAA85324.1; -.  
DR InterPro; IPR000883; COX1.  
DR Pfam; PF00115; COX1; 1.  
DR PRINTS; PR01165; CYCOXIDASE1.  
DR PROSITE; PS00077; COX1; UNKNOWN\_1.  
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
FT NON\_TER 1 275  
SQ SEQUENCE 275 AA; 30298 MW; 20B1917902CB529D CRC64;

Query Match 48.0%; Score 47; DB 8; Length 275;  
Best Local Similarity 50.0%; Pred. No. 9.5;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FQANCGIDFIIF 12  
|:|:|:|:|:|  
Db 36 FSSSCGVDFLMF 47

## RESULT 4

Q9EV68  
ID Q9EV68 PRELIMINARY; PRT; 491 AA.  
AC Q9EV68;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Putative NADH-ubiquinone oxidoreductase subunit.  
GN NUOM3.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RM41;  
RA Putnoky P., Jady B., Chellapilla K.P., Barta F., Kiss E.;  
RT "Sinorhizobium meliloti carries two sets of nuo genes."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; AJ245399; CAC14150.1; -.  
DR InterPro; IPR003918; NADHub\_oxred4.  
DR InterPro; IPR001750; Oxidored\_q1.  
DR Pfam; PF00361; oxidored\_q1; 1.  
DR PRINTS; PR01437; NUOXDRDTASE4.  
KW NAD; Oxidoreductase; Transmembrane; Ubiquinone.  
SQ SEQUENCE 491 AA; 52714 MW; A2C62574220C4EEA CRC64;

Query Match 48.0%; Score 47; DB 2; Length 491;  
Best Local Similarity 41.7%; Pred. No. 16;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 5 CGIDFIIFWIFW 16  
|:|:|:|:|:|  
Db 127 CALDLFLFYVFW 138

## RESULT 5

Q92YN6  
ID Q92YN6 PRELIMINARY; PRT; 491 AA.





DR PROSITE; PS50262; G\_PROTEIN\_RECEP\_FL\_2; 1.  
KW Receptor.  
SQ SEQUENCE 310 AA; 34917 MW; 452A1D723081781D CRC64;  
  
Query Match 46.9%; Score 46; DB 11; Length 310;  
Best Local Similarity 50.0%; Pred. No. 15;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 QANCGIDFIIFWIF 15  
Db 17 QLSAGIEMFLWIF 30  
  
RESULT 9  
Q953M3 PRELIMINARY; PRT; 530 AA.  
AC Q953M3;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).  
GN COX1.  
OS Echinococcus granulosus.  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;  
OC Cyclophyllidae; Taeniidae; Echinococcus.  
OX NCBI\_TaxID=6210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Le T.H., Dai T.H.N., Blair D., McManus D.P.;  
RT "Complete mitochondrial genome of Echinococcus granulosus (G4 genotype).";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.  
CC -!- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
DR EMBL; AF346403; AAK82350.1; -.  
DR InterPro; IPR000883; COX1.  
DR Pfam; PF00115; COX1; 1.  
DR PRINTS; PR01165; CYCOXIDASE1.  
DR PROSITE; PS00077; COX1; UNKNOWN\_1.  
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;  
KW Respiratory chain; Transmembrane.  
FT NON\_TER 1  
SQ SEQUENCE 530 AA; 59632 MW; 2E5FCA534DFD1771 CRC64;  
  
Query Match 46.9%; Score 46; DB 8; Length 530;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 QANCGIDFIIF 12  
Db 133 FSSGGGVDFLMF 144  
  
RESULT 10  
Q9RPH1 PRELIMINARY; PRT; 3223 AA.  
AC Q9RPH1;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE EHEC factor for adherence.  
GN EFAL.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E45035;  
RX MEDLINE=20117988; PubMed=10652089;  
RA Nicholls L., Grant T.H., Robins-Browne R.M.;  
RT "Identification of a novel genetic locus that is required for in vitro adhesion of a clinical isolate of enterohaemorrhagic Escherichia coli to epithelial cells.";  
RL Mol. Microbiol. 35:275-288(2000).  
DR EMBL; AF159462; AAD49229.2; -.  
DR InterPro; IPR001917; NHtransf\_2.  
DR InterPro; IPR000169; SHprot\_acsite.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; UNKNOWN\_1.  
DR PROSITE; PS00639; THIOI\_PROTEASE\_HIS; UNKNOWN\_1.  
SQ SEQUENCE 3223 AA; 365949 MW; CAFD59F69242D07A CRC64;  
  
Query Match 46.9%; Score 46; DB 2; Length 3223;  
Best Local Similarity 43.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 FOANCGIDFIIFWIF 16  
Db 2163 YQANTEIEYLSNWIVW 2178  
  
RESULT 11  
Q9RM48 PRELIMINARY; PRT; 3223 AA.  
AC Q9RM48;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Lymphostatin.  
GN LIFA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E2348/69;  
RX MEDLINE=20187489; PubMed=10722613;  
RA Klaproth J.M.A., Scaletsky I.C.A., McNamara B.P., Lai L.C.,  
RA Malstrom C., James S.P., Donnenberg M.S.;  
RT "A large toxin from pathogenic Escherichia coli strains that inhibits lymphocyte activation.";  
RL Infect. Immun. 68:2148-2155(2000).  
DR EMBL; AJ133705; CAB55629.1; -.  
DR InterPro; IPR001917; NHtransf\_2.  
DR InterPro; IPR000169; SHprot\_acsite.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; UNKNOWN\_1.  
DR PROSITE; PS00639; THIOI\_PROTEASE\_HIS; UNKNOWN\_1.  
SQ SEQUENCE 3223 AA; 365963 MW; 03E45ECDED7938C1 CRC64;  
  
Query Match 46.9%; Score 46; DB 2; Length 3223;  
Best Local Similarity 43.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 FOANCGIDFIIFWIF 16  
Db 2163 YQANTEIEYLSNWIVW 2178  
  
RESULT 12  
Q8VQR2 PRELIMINARY; PRT; 3223 AA.  
ID Q8VQR2  
AC Q8VQR2;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Efa1.  
GN EFa1.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=83/39;  
RX Tauschek M., Strugnell R.A., Robins-Browne R.M.;  
RA "Characterization of the LEE pathogenicity islands of rabbit  
RT enteropathogenic Escherichia coli.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF453441; AAL57562.1; -.  
DR InterPro; IPR001917; NHtransf\_2.  
DR InterPro; IPR00169; SHprot\_acsite.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; UNKNOWN\_1.  
DR PROSITE; PS00639; THIOLESTERASE\_HIS; UNKNOWN\_1.  
SQ SEQUENCE 3223 AA; 365790 MW; E97D10B98FADE658 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 3223;  
Best Local Similarity 43.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFW 16  
:|||||:|:|  
Db 2163 YQANTEIEYLSNWIVW 2178

## RESULT 13

Q8VNR6 PRELIMINARY; PRT; 3223 AA.  
AC Q8VNR6;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Efa1-Lifa-Tox protein.  
GN EFa1-LIFA-TOX.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=413/89-1;  
RA Benkel P., Chakraborty T.;  
RT "Genetic organisation and sequence of the LEE II locus in Shiga toxin-  
RT producing Escherichia coli.";  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ277443; CAC81883.1; -.  
DR InterPro; IPR001917; NHtransf\_2.  
DR InterPro; IPR00169; SHprot\_acsite.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; UNKNOWN\_1.  
DR PROSITE; PS00639; THIOLESTERASE\_HIS; UNKNOWN\_1.  
SQ SEQUENCE 3223 AA; 365666 MW; 28EBB4374FAB8FC7 CRC64;

Query Match 46.9%; Score 46; DB 2; Length 3223;  
Best Local Similarity 43.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFW 16  
:|||||:|:|  
Db 2163 YQANTEIEYLSNWIVW 2178

## RESULT 14

Q968Z8 PRELIMINARY; PRT; 269 AA.  
ID Q968Z8  
AC Q968Z8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Ookinete surface protein Pos28-2.  
GN POS28-2.  
OS Plasmodium ovale (malaria parasite P. ovale).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIGERIAN I/CDC;  
RX MEDLINE=21192570; PubMed=11295191;  
RA Tachinbana M., Tsuboi T., Templeton T.J., Kaneko O., Torii M.;  
RT "Presence of three distinct ookinete surface protein genes, Pos25,  
RT Pos28-1, and Pos28-2, in Plasmodium ovale.";  
RL Mol. Biochem. Parasitol. 113:341-344(2001).  
DR EMBL; AB051633; BAB43950.1; -.  
DR InterPro; IPR000561; EGF-like.  
DR SMART; SM00181; EGF; 3.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.  
SQ SEQUENCE 269 AA; 29928 MW; 7A949BB42F036EEB CRC64;

Query Match 45.9%; Score 45; DB 5; Length 269;  
Best Local Similarity 46.2%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 FOANCGIDFIIFW 13  
|:|:|:|:|  
Db 92 FKCNCTRNIVFW 104

## RESULT 15

Q91ZB6 PRELIMINARY; PRT; 319 AA.  
ID Q91ZB6;  
AC Q91ZB6;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE G protein-coupled receptor (Fragment).  
GN MRGF.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=21435808; PubMed=11551509;  
RA Dong X., Han S.-K., Zylka M.J., Simon M.I., Anderson D.J.;  
RT "A Diverse Family of GPCRs Expressed in Specific Subsets of  
RT Nociceptive Somatosensory Neurons.";  
RL Cell 106:619-632(2001).  
DR EMBL; AY042211; AAK91802.1; -.  
DR InterPro; IPR000276; GPCR\_Rhodpsn.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; UNKNOWN\_1.  
DR PROSITE; PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
KW Receptor.  
FT NON\_TER  
SQ SEQUENCE 319 AA; 35955 MW; D1860538BA7A965C CRC64;

Query Match 45.9%; Score 45; DB 11; Length 319;  
Best Local Similarity 60.0%; Pred. No. 23;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 GIDFIIFWIF 15  
|||:|:|:|  
Db 232 GIDWFLFWVF 241

Search completed: July 9, 2003, 15:06:33  
Job time : 10.2212 secs

